

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:11:16 ; Search time 825.255 Seconds  
(without alignments)  
1233.024 Million cell updates/sec

Title: US-10-777-131a-1

Perfect score: 21

Sequence: 1 tccgagatgatcttactcc 21

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hcg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sbs: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	6	AX770103 Sequence
2	21	100.0	823	6	AX770121 Sequence
3	21	100.0	3002	6	CQ880370 Sequence
4	21	100.0	40056	1	MSGY348
5	21	100.0	110000	1	AE000516_12
6	21	100.0	327650	1	BX248337 Mycobacte
7	21	100.0	349306	1	BX842575 Mycobacte
8	17.8	84.8	144190	2	AP004112 Oryza sat
9	17.8	84.8	166453	8	AP005649
10	17.8	84.8	175567	9	AC079947 Homo sapi
11	17.8	84.8	195744	2	AC092002 Homo sapi
12	17.8	84.8	247746	2	AC112285 Rattus no
13	17.8	84.8	291544	2	AC128496 Rattus no
14	17.8	84.8	320731	2	AC149573 Medicago
15	17.4	82.9	1286	8	ATU90428 Arabidops
16	17.4	82.9	109375	8	AC003672 Arabidops
17	17.4	82.9	152709	2	AC149508 Strongylo
18	17	81.0	472	10	BC025166 Mus muscu
19	17	81.0	4578	10	AF104410 Mus muscu

C	20	17	81.0	11013	1	AE011443	AE011443 Leptospir
C	21	17	81.0	109431	9	AC016596	AC016596 Homo sapi
C	22	17	81.0	138905	2	AC131507	AC131507 Strongylo
C	23	17	81.0	144228	9	AC068599	AC068599 Homo sapi
C	24	17	81.0	223276	10	AL593853	AL593853 Mouse DNA
C	25	17	81.0	300478	1	AE017201	AE017201 Lactobaci
C	26	17	81.0	301508	1	AE017292	AE017292 Leptospir
C	27	17	81.0	349980	6	AX926711	AX926711 Sequence
C	28	17	81.0	349980	6	AX926712	AX926712 Sequence
C	29	16.8	80.0	519	3	AY488869	AY488869 Plasmodu
C	30	16.8	80.0	519	3	AY488871	AY488871 Plasmodu
C	31	16.8	80.0	4308	6	AR528296	AR528296 Sequence
C	32	16.8	80.0	10358	1	AE013482	AE013482 Methanosa
C	33	16.8	80.0	28149	8	YSC19931	Y0162 Saccharomyc
C	34	16.8	80.0	90132	8	AC084165	AC084165 Arabidops
C	35	16.8	80.0	110000	1	CR522870_29	Continuation (30 o
C	36	16.8	80.0	117919	8	AP003220	AP003220 Oryza sat
C	37	16.8	80.0	131239	9	HS667H12	AL035414 Human DNA
C	38	16.8	80.0	133060	2	AC138555	AC138555 Human DNA
C	39	16.8	80.0	137678	9	HS879X22	AL034351 Human DNA
C	40	16.8	80.0	164138	2	AC109432	AC109432 Rattus no
C	41	16.8	80.0	166143	2	AL356960	AL356960 Homo sapi
C	42	16.8	80.0	169486	8	AP005648	AP005648 Oryza sat
C	43	16.8	80.0	208155	9	AC124254	AC124254 Homo sapi
C	44	16.8	80.0	228736	2	AC107161	AC107161 Rattus no
C	45	16.4	78.1	3292	1	AY038363	AY038363 Shewanell

#### ALIGNMENTS

RESULT 1	AX770103	Sequence 1 from Patent WO03016562.	21 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770103					
DEFINITION	Sequence 1 from Patent WO03016562.					
ACCESSION	AX770103					
VERSION	AX770103.1	GI:32437681				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
ORIGIN						
Query Match						
Best Local Similarity	100.0%;	Score 21;	DB 6;	Length 21;		
Matches	21;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	TCCGATGATGATTTTACCTCC	21			
Db	1	TCCGATGATGATTTTACCTCC	21			
RESULT 2	AX770121	Sequence 19 from Patent WO03016562.	823 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770121					
DEFINITION	Sequence 19 from Patent WO03016562.					
ACCESSION	AX770121					
VERSION	AX770121.1	GI:32437699				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
ORIGIN						
Query Match						
Best Local Similarity	100.0%;	Score 21;	DB 6;	Length 21;		
Matches	21;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	TCCGATGATGATTTTACCTCC	21			
Db	1	TCCGATGATGATTTTACCTCC	21			

**REFERENCE**  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.  
**AUTHORS**  
 1 Glacquel, B.  
**TITLE**  
 Compositions and methods for detecting multidrug resistant strains  
 of M. tuberculosis having mutations in genes of the mutR family  
**JOURNAL**  
 Patent: WO 03016562-A 19 27-FEB-2003,  
 INSTITUT PASTEUR (FR)  
**FEATURES**  
 Location/Qualifiers  
 source 1..823  
 /organism="Mycobacterium tuberculosis"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:1773"  
**ORIGIN**  
 Query Match 100.0%; Score 21; DB 6; Length 823;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
**QY**  
 1 TCCGGATGATGATTACCTCC 21  
 |||||  
 5 TCCGGATGATGATTACCTCC 25  
**RESULT 3**  
 CO880370 3002 bp DNA linear PAT 11-OCT-2004  
**LOCUS**  
 CO880370  
**DEFINITION**  
 Sequence 1 from Patent WO2004083459.  
**ACCESSION**  
 CO880370.1 GI:54034136  
**VERSION**  
 CO880370.1  
**KEYWORDS**  
 Mycobacterium tuberculosis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.  
**SOURCE**  
 1  
 Bange, F.C.  
 Method and kit for a specific detection of m.tuberculosis  
 Patent: WO 2004083459-A 1 30-SEP-2004;  
 Artus Gesellschaft Fuer Molekularbiologische Diagnostik un d  
 Entwicklung MBH. (DE)  
**FEATURES**  
 Location/Qualifiers  
 source 1..3002  
 /organism="Mycobacterium tuberculosis"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:1773"  
**ORIGIN**  
 Query Match 100.0%; Score 21; DB 6; Length 3002;  
 Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
**QY**  
 1 TCCGGATGATGATTACCTCC 21  
 |||||  
 788 TCCGGATGATGATTACCTCC 808  
**Db**  
 788 TCCGGATGATGATTACCTCC 808  
**RESULT 4**  
 MSGY348 40056 bp DNA linear BCT 10-DEC-1996  
**LOCUS**  
 MSGY348  
**DEFINITION**  
 Mycobacterium tuberculosis sequence from clone y348.  
**ACCESSION**  
 ADO000020  
 ADO000020.1 GI:1717739  
**VERSION**  
 ADO000020.1  
**KEYWORDS**  
 Mycobacterium tuberculosis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.  
**SOURCE**  
 1  
 Du, L.  
**REFERENCE**  
 1 (bases 1 to 40056)  
 Du, L.

**TITLE**  
 Direct Submission  
**JOURNAL**  
 Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100  
 Beaver Street, Waltham, MA, USA, 02154 du@eric.com  
**COMMENT**  
 GSDS:8:1004716.  
**FEATURES**  
 Location/Qualifiers  
 source 1..40056  
 /organism="Mycobacterium tuberculosis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1773"  
 /clone="y348"  
**ORIGIN**  
 Query Match 100.0%; Score 21; DB 1; Length 40056;  
 Best Local Similarity 100.0%; Pred. No. 3.5;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
**QY**  
 1 TCCGGATGATGATTACCTCC 21  
 |||||  
 17919 TCCGGATGATGATTACCTCC 17939  
**Db**  
 17919 TCCGGATGATGATTACCTCC 17939  
**RESULT 5**  
 AE000516\_12  
**WPCOMMENT**  
 Sequence split into 44 fragments LOCUS AE000516 Accession AE000516  
 Fragment Name Begin End  
 AE000516\_00 1 110000  
 AE000516\_01 100001 210000  
 AE000516\_02 200001 310000  
 AE000516\_03 300001 410000  
 AE000516\_04 400001 510000  
 AE000516\_05 500001 610000  
 AE000516\_06 600001 710000  
 AE000516\_07 700001 810000  
 AE000516\_08 800001 910000  
 AE000516\_09 900001 1010000  
 AE000516\_10 1000001 1110000  
 AE000516\_11 1100001 1210000  
 AE000516\_12 1200001 1310000  
 AE000516\_13 1300001 1410000  
 AE000516\_14 1400001 1510000  
 AE000516\_15 1500001 1610000  
 AE000516\_16 1600001 1710000  
 AE000516\_17 1700001 1810000  
 AE000516\_18 1800001 1910000  
 AE000516\_19 1900001 2010000  
 AE000516\_20 2000001 2110000  
 AE000516\_21 2100001 2210000  
 AE000516\_22 2200001 2310000  
 AE000516\_23 2300001 2410000  
 AE000516\_24 2400001 2510000  
 AE000516\_25 2500001 2610000  
 AE000516\_26 2600001 2710000  
 AE000516\_27 2700001 2810000  
 AE000516\_28 2800001 2910000  
 AE000516\_29 2900001 3010000  
 AE000516\_30 3000001 3110000  
 AE000516\_31 3100001 3210000  
 AE000516\_32 3200001 3310000  
 AE000516\_33 3300001 3410000  
 AE000516\_34 3400001 3510000  
 AE000516\_35 3500001 3610000  
 AE000516\_36 3600001 3710000  
 AE000516\_37 3700001 3810000  
 AE000516\_38 3800001 3910000  
 AE000516\_39 3900001 4010000  
 AE000516\_40 4000001 4110000  
 AE000516\_41 4100001 4210000  
 AE000516\_42 4200001 4310000  
 AE000516\_43 4300001 4403837  
 Continuation (13 of 44) of AE000516 from base 1200001 (AE000516 Mycobacterium tuberculosis  
 Query Match 100.0%; Score 21; DB 1; Length 110000;

Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OR 1 TCCGATGATGATTTACTCC 21  
|||||  
|||||  
Db 85887 TCCGATGATGATTTACTCC 85907

RESULT 6  
BX248337 327650 bp DNA linear BCT 06-MAY-2004  
LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment  
DEFINITION 4/14.  
ACCESSION BX248337 BX248337  
VERSION BX248337.1 GI:31617663  
KEYWORDS complete genome.  
SOURCE Mycobacterium bovis AF2122/97  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE 1  
AUTHORS Garnier,T., Biglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,  
Pryor,M., Duchoy,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,  
Harris,B., Aikin,R., Doggett,J., Mayes,R., Keating,L.,  
Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and  
Hewison,G.  
The complete genome sequence of Mycobacterium bovis  
Unpublished  
2 (bases 1 to 327650)  
TITLE Direct Submission  
AUTHORS Garnier,T.  
JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire  
Bacterieme Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex  
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the  
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary  
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,  
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK. Pr4 Annotation, Genopole, Institut  
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
Unite de Genetique Moleculaire Bacterieme, Institut Pasteur, 28  
rue du Docteur Roux, 75724 Paris Cedex 15, France

FEATURES  
source 1. 327650  
/organism="Mycobacterium bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:233413"  
698..2170  
/locus\_tag="Mb0912"  
698..2170  
/locus\_tag="Mb0912"  
698..2170  
/note="Mb0912", len: 490 aa. Equivalent to Rv0888, len:  
490 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 490 aa overlap). Probable exported  
protein. Equivalent to AAK45157.1 from Mycobacterium  
tuberculosis strain CDC1551 (507 aa) but shorter 17 aa.  
Contains possible N-terminal signal sequence."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE EXPORTED PROTEIN"  
/protein\_id="CAD93774.1"  
/db\_xref="GI:31617664"  
/db\_xref="GOA:Q10549"  
/db\_xref="UniProt/Swiss-Prot:Q10549"  
/translation="MDYAKRIGGVGALAVTVGAAYTTNIGSAAPTPSSSTDSBP  
VDACSPPGSSASLSAIPGASVPOVAVDPCGIPDDILNALIDFLAAVNGVLPIT  
ENRTVPANPQGSVSEGGVGVPRFDCDDPGNMTFAVRERGA PGGQHVAVGVVDR  
TASRTADRGVGTDPFVSVDSTDSIHVHGAGVYCFRFGHCDVATVTVFVNTPT  
DTISGDSMLTYNAGIPPLSSAILPRFFTKTIGKRLNAYVYVANOEDPAHYQFLI  
KSKMPSOTPEPEPTLMPIGVPSDGLNTLSERKVGQLDQTVTECTSDCLTKEP  
TYSQRLPGGVTVVNLHTVTGGPTTNALAAVANYIQNSAGRAVIVTGDNNAY  
TSDQSALQFAOVNGLTDNAVQVHEHGTTPPEAFATCWGNCBILDKIFYRSGGVTL

gene  
CDS

CAVSYGNAPKRFNSKCEPLSDHSFAVGHVYADVNAVNR"  
complement(2205..3326)  
/gene="cItA"  
/locus\_tag="Mb0913c"  
complement(2205..3326)  
/gene="cItA"  
/locus\_tag="Mb0913c"  
/EC\_number="2.3.3.1"  
/note="Mb0913c, cItA, len: 373 aa. Equivalent to Rv0889c,  
len: 373 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 373 aa overlap). Probable cItA  
alternate gene name: gItA, citrate synthase 2 (EC  
4.1.3.7), highly similar to others e.g.  
CA85899.1|A435998 putative citrate synthase from  
Streptomyces coelicolor (367 aa); P3119|CISY\_BACSU  
citrate synthase II from Bacillus subtilis (366 aa), FASTA  
scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa  
overlap); etc. Also similar to Rv0896|MYCY31.24 from  
Mycobacterium tuberculosis (29.2% identity in 274 aa  
overlap) and Rv1131. Contains P500480 Citrate synthase  
signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE CITRATE SYNTHASE II CITA"  
/protein\_id="CAD93774.1"  
/db\_xref="GI:31617665"  
/db\_xref="GOA:Q10529"  
/db\_xref="UniProt/Swiss-Prot:Q10529"  
/translation="MTVPENFVGLGVGVAFTTEIEAPDKGALRYGVDIRDLVS  
QRYVFGVMAVLVDGNGSGLPAPPEPLPIHSGDVIVDQAGLAMIPIGVAPLD  
IDPATRQGLARASVMAISYVAOSARGIYOPAVORTIDECTARFMTGQEPDP  
RHTETADYVVAASHGMSASTPARIASGTADVAALSGAIGAMGPRVCGAPRV  
LPMIDVERAGDARSVKGILDRCEKLMGCRHYRYRADPPRVLRAAEKLGARVE  
VIAAVEQALASELRERRPDRIETNVEFAVVDFAVPPANMPAMFTCGRTGWCA  
HVEQKRIKGLVYPSAISYVGPSPESVDGWERVLTJA"  
complement(3413..6061)  
/locus\_tag="Mb0914c"  
complement(3413..6061)  
/locus\_tag="Mb0914c"  
/note="Mb0914c", len: 882 aa. Equivalent to Rv0890c,  
len: 882 aa, from Mycobacterium tuberculosis strain H37Rv,  
(99.9% identity in 882 aa overlap). Probable  
transcriptional regulatory protein, luxR family, highly  
similar (but shorter 238 aa in N-terminus) to  
NP\_302202.1|NC\_002677 possible transcriptional regulator  
from Mycobacterium leprae (1106 aa). Also highly similar  
(generally in part) to others e.g. T50568 probable  
multi-domain regulatory protein from Streptomyces  
coelicolor (1334 aa); P10957|NARL\_ECOLI nitrate/nitrite  
response regulator protein from Escherichia coli (216 aa),  
FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 99  
aa overlap); etc. Also highly similar to others from  
Mycobacterium tuberculosis e.g. MTCY02810.22, MTVO08.44,  
MTY036.21, and MTCY11.24. Contains P500017 ATP/GTP-binding  
site motif A (P-loop), P500622 Bacterial regulatory  
proteins, luxR family signature, and probable helix-turn  
helix motif from aa 836 to 857 (score 1559, +4.50 SD).  
BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL  
REGULATORS."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN  
(PROBABLY LUXR-FAMILY)"  
/protein\_id="CAD93775.1"  
/db\_xref="GI:31617666"  
/db\_xref="UniProt/Swiss-Prot:P59969"  
/translation="MRALLQONRLVTLGCGVGTSLAIOIASASELBDGLCFVDLA  
PIRESGLVATARAVALGPDPQGRSTDSLRFRFGKNTMLAVDNCNGLHLDACAAVY  
ELGAGCELTILATSEPTICAGETIRMSVSTIETPAVGLDRAAGVAGPCTIAH  
NAAAGETICRLNGIPLAIFPAARVRSMSPLEIADGLDFRILASRGVAGVQROOT  
LRASIDSHALLTETEDILFRRLAPFVGGEDLAIVRAVVAAGSDIDPSVVDQLTLD  
KSVLVADDCCGRRRLVETVRYVALAEKGDSEADVAHARHLYTTALASLNPAN  
DHQRLVARAETEDINLAAPAWSRNENHITAEQLQLASLQIPWFRGRLHREGLSFNS

gene  
CDS

ILBQRPHRLAVTAFAVBARALADKAMSTLWATSPVATITLAPAOALIMARVQGP  
 ALVALRAITAGCGSGNYAAEAAPFAPATATDLARLIDKRTILCOILYKRGCTCISGPP  
 NALTAABEECDLDDTIGDRFVSRHCSLWLSLAQMGNTLEALSREYITAEKASBN  
 DVPLKVLGLYTAQVLAFCGASAMHAIAGCTAATBELGVYOGIYAAMTYALAAAG  
 DVTALFASDAARPIILRAQPOVYTHVLAQMLAAGSDLAARQFANDVADTGNH  
 RMVALITRARVATARGSEPELARDHAALHAGALHITXQGMPEMLIAGAGVSH  
 SEGRILIGAAALROOTROYEFKIDMGVQAVTALBEMDEDFDAMAGALSTDS  
 EALVYKRGGRGRRPARAGMSLTPTERDVRLVSBGLSNDIAKRLFVSPRYQTHL  
 THVYKGLASRVOLVDEBARRGSPS"  
 complement (6063. .6920)  
 /locus\_tag="Mb0915c"  
 /complement (6063. .6920)  
 /locus\_tag="Mb0915c"  
 /note="Mb0915c", len: 285 aa. Equivalent to Rv0891c,  
 len: 285 aa, from Mycobacterium tuberculosis strain H37Rv,  
 (99.6% identity in 285 aa overlap). Possible  
 transcriptional regulator, highly similar in N-terminus to  
 NP\_302202.1|NC\_002677 possible transcriptional regulator  
 from Mycobacterium leprae (1106 aa). Also highly similar  
 to several Mycobacterium tuberculosis putative  
 transcriptional regulators e.g. Q1102|MT02B10\_22  
 PROBABLE TRANSCRIPTIONAL REGULATOR PROTEIN (1159 aa),  
 PASTA scores: opt: 702, E(): 8.3e-40, (50.6% identity in  
 247 aa overlap); MT0036\_21; MT0008\_44; MT002B10\_23. Also  
 shows similarity with several adenylate cyclases and  
 hydrolases from other organisms."  
 /codon\_start=1  
 /transl\_table=1  
 /product="POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN"  
 /protein\_id="CAD93776.1"  
 /db\_xref="GI:31617667"  
 /db\_xref="UniProt/Swiss-Prot:P59970"  
 /translation="MLRPAVYNSLPRLPIDHATLRGDHPPTCAKCYARIGATLSSG  
 LIDLVHSRCYAAPPDVGRCSFPPRRVRLRANGGLDSRLPPTGYTLADVEEST  
 HLWMLCPEDMAITAIALHDVTSEALITNGGVQPKRYEGSFVAFTASPAALCALDD  
 LQRLSLPRLRLRIGLITGEVQRLDELVYPIINTARLDIARHGQVVLSPATDILVYV  
 GRLPADAMLVLDGRHPLRGLRPRVWMLCPDIRKEKPPPLRTAKSPSTSLPAQFTTH  
 FGRRAOIS"  
 7318. .8805  
 /locus\_tag="Mb0916"  
 7318. .8805  
 /locus\_tag="Mb0916"  
 /EC\_number="1.14.-.-"  
 /note="Mb0916", len: 495 aa. Equivalent to Rv0892, len:  
 495 aa, from Mycobacterium tuberculosis strain H37Rv,  
 (100.0% identity in 495 aa overlap). Probable  
 monooxygenase (EC 1.14.-.-), highly similar to others e.g.  
 NP\_250787.1|NC\_002516 probable flavin-binding  
 monooxygenase from Pseudomonas aeruginosa (491 aa);  
 CAB59668.1|AL132674 monooxygenase from Streptococcus  
 coelicolor (519 aa); P12015|CWMO\_ACIS cyclohexanone  
 monooxygenase from Acinetobacter sp. (542 aa), PASTA  
 scores: opt: 489, E(): 6.8e-26, (30.3% identity in 492 aa  
 overlap); etc. Also highly similar to Rv0565c, Rv3684c,  
 Rv31083, etc from Mycobacterium tuberculosis. Has  
 hydrophobic stretch at N-terminus."  
 /codon\_start=1  
 /transl\_table=1  
 /product="PROBABLE MONOOXYGENASE"  
 /protein\_id="CAD93777.1"  
 /db\_xref="GI:31617668"  
 /db\_xref="GOA:O10532"  
 /db\_xref="UniProt/Swiss-Prot:O10532"  
 /translation="MTGRCPTVAAYGAGSMCAVITLLSGITDVCIYERADVDVGGT

RESULT 7				
LOCUS	BX842575	349306 bp	DNA	linear BCT 10-JUN-2004
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 4/13.			
ACCESSION	BX842575 AL010186 AL021006 AL021897 AL021999 AL123456 Z79700 Z92539 Z93777 Z94752 Z95209 Z95210 Z95584 Z95585 Z98260			
VERSION	BX842575.1 GI:41353619			
KEYWORDS	complete genome,			
SOURCE	Mycobacterium tuberculosis H37Rv			
ORGANISM	Mycobacterium tuberculosis H37Rv Bacteria; Actinobacteria; Actinomycetia; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.			
REFERENCE	1			
AUTHORS	Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E., Tearl,A.F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Comor,R., Davies,R., Devlin,K., Fellwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., Mclean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Radarsman,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., Squires,S., Squires,J., Sulston,P.E., Taylor,K., Whitehead,S. and Barrell,B.G. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence Nature 393 (6685), 537-544 (1998)			
JOURNAL	Nature 393 (6685), 537-544 (1998)			
PUBMED	98295887			
REFERENCE	2			
AUTHORS	Canus,J.C., Pryor,M.J., Medigue,C. and Cole,S.T. Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)			
JOURNAL	Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)			
PUBMED	22255591			
REFERENCE	3			
AUTHORS	(bases 1 to 349306) Parkhill,J. Direct Submission Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On or before Jan 28, 2004 this sequence version replaced gi:3261499, gi:3256022, gi:3261558, gi:3261628, gi:3261714, gi:3261728, gi:3261731, gi:3261750, gi:3261774, gi:3261787, gi:3261826. Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/). Location/Qualifiers 1..349306 /organism="Mycobacterium tuberculosis H37Rv" /mol_type="genomic DNA" /strain="H37Rv" /db_xref="taxon:83332" complement(53..844) /locus_tag="RV0927c" complement(53..844) /locus_tag="RV0927c" /EC_number="1.-.-.-" /function="UNKNOWN; POSSIBLY INVOLVED IN CELLULAR METABOLISM."" /note="RV0927c, (MTYC21C12_21c), len: 263 aa. Probable short-chain dehydrogenase/reductase (BC 1.-.-.-), similar to various dehydrogenases/dehydrogenases, notably 7-alpha-hydroxyesteroid dehydrogenases and glucose 1-dehydrogenases e.g. p25529 DHHA_ECOLI 7-alpha-hydroxyesteroid dehydrogenase from Escherichia coli (255 aa), FASTA scores: opt: 551, E(1): 1e-26, (39..54 identity in 248 aa overlap); NP_252778.1 NC_002516			
COMMENT				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				



probable sort-chain dehydrogenase from *Pseudomonas aeruginosa* (253 aa); AAC44307.1|U59433.3-kecacyl-acyl carrier protein reductase from *Bacillus subtilis* (246 aa); etc. Also similar to other dehydrogenases from *Mycobacterium tuberculosis* e.g. MTCY09P9.36, E|1.4e-18; MTCY369.14, E|1.8e-17; MTCY02B10.14, E|2.5e-14; MTCY09F9.23C, E|1.5e-13; MTCY03C7.07, E|1.9e-13. Contains P500061 Short-chain dehydrogenases/reductases family signature, and P50017 ATP/GTP-binding site motif A (P-loop). BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDOUCTASES (SDR) FAMILY. "

//codon\_start=1  
//transl\_table=11  
//product="PROBABLE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE"  
//protein\_id="CAB84497.1"  
//db\_xref="GI:2078076"  
//translation="MLIMIDRGAIVTGGGRLGAGALALAPAQAGADVLASRTS  
SELDVAEQLRARGRRMYTADLAPETVALAQAQVAFKDLIVVNNGTGMPPT  
LITSTQALHDYAFNFGVGLHATVLAAGSLIMHSGGGSVINISTGRLAARFALY  
GLTAKALHTATRLIALDLCGRVAVNALIAGSLITLSEVANDETRAPMGAIFLAR  
LGDVDIDAAAVYASAGSRFLGTLEVDGSLTPPNLDLPPLD"  
//complement(113..136)  
//locus\_tag="Rv0927c"  
//name="P500017 ATP/GTP-binding site motif A (P-loop)  
//complement(320..406)  
//locus\_tag="Rv0927c"  
//note="P500061 Short-chain alcohol dehydrogenase family signature"  
1116..2228  
//gene="pncS3"  
//locus\_tag="Rv0928"  
1116..2228  
//gene="pncS3"  
//locus\_tag="Rv0928"  
//function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT). THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT."  
//standard\_name="phoS2"  
//note="Rv0928. (MTCY21C12.22), len: 370 aa. pncS3 (previously known as phoS2), phosphate-binding lipoprotein component of inorganic phosphate transport system (see citations below), highly similar to others from *Mycobacterium leprae* e.g. Q50099|PSTR3|PHOS1 phosphate-binding protein 3 precursor (328 aa), FASTA scores: opt: 1772, E|: 0.179.6% identity in 328 aa overlap), and highly similar to others e.g. AAF74819.1|AF137360.1|AF137360 periplasmic phosphate permease from *Mycobacterium avium* (369 aa). Also highly similar to Rv0933C|MTCY0890.07|pncS2 PHOSPHATE-BINDING PERIPLASMIC LIPOPROTEIN (370 aa); and Rv0934|pncS1 PHOSPHATE-BINDING PERIPLASMIC LIPOPROTEIN (374 aa) from *Mycobacterium tuberculosis* (*Mycobacterium tuberculosis* seems to have three pncS-like proteins, others being Rv0932c and Rv0934c). Contains lipoprotein signature (P500013) at N-terminus. BELONGS TO FAMILY OF PHOSPHATE RECEPTORS FOR BACTERIAL ABC-TYPE LIPOPROTEIN TRANSPORTERS. "

//codon\_start=1  
//evidence=experimental  
//transl\_table=11  
//product="PERIPLASMIC PHOSPHATE-BINDING LIPOPROTEIN P5T53 (PBP-3) (P5T53) (PHOS1)"  
//protein\_id="CAE55337.1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYRKRRLRSATVANA1"  
//db\_xref="GI:11353620"  
//translation="MKLRKGAIVGLAALVLSACGINDNVTCGATLGOAAKDY  
CGKTLKASGTSQAAAMRFVFNVEQACPGQTLVYNGSAGISFNENQPDGCG  
SDVLSDEDAALARRCGSAMLLPVPRBIAVITFNYSVSLINDGTGLKINSGI  
TQNNPVALQALNDFPLRGRHIVRSPDSGTTTNRVYLDAASNGWGAACKSRQ  
GVGGAGANDGSSAAKNNPSSITNNSSFMQDHLTNANVLSAGDPAVLTIDSV  
QQTLAGATISGVNDVLVDTSYR

LOCUS	DEFINITION	ACCESION	misc_feature
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	1152..1184 /gene="pbcS3" /locus_tag="Rv0928" /note="P500013 Prokaryotic membrane lipoprotein lipid attachment site"
LOCUS	DEFINITION	ACCESION	gene
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	2241..3215 /gene="pbcC2" /locus_tag="Rv0929" 2241..3215 /gene="pbcC2" /locus_tag="Rv0929" /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT." /note="Rv0929, (MTCY21C12.23), len: 324 aa. pbcC2, phosphate-transport integral membrane ABC transporter (see citations below), highly similar to others e.g. NP_302394.1 NC_002677 membrane-bound component of phosphate transport from Mycobacterium leprae (319 aa); CAB86474.1 AF353816 phosphate ABC transport system permease protein from Streptomyces coelicolor (336 aa); NP_290359.1 NC_002655 high-affinity phosphate-specific transport system (cytoplasmic membrane component) from Escherichia coli strain 0157:H7 (319 aa); etc. Also similar to Rv935 MTCY08D9.04 PSTC1 PROBABLY TRANSMEMBRANE ABC TRANSPORTER COMPONENT OF PHOSPHATE UPTAKE SYSTEM from Mycobacterium tuberculosis (338 aa). Contains binding-protein-dependent transport systems inner membrane component signature (P500402)."
LOCUS	DEFINITION	ACCESION	misc_feature
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	2844..2930 /gene="pbcC2" /locus_tag="Rv0929" /note="P500402 Binding-protein-dependent transport systems inner membrane comp signature"
LOCUS	DEFINITION	ACCESION	gene
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	3212..4138 /gene="pbcA1" /locus_tag="Rv0930" 3212..4138 /gene="pbcA1" /locus_tag="Rv0930" /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE."
LOCUS	DEFINITION	ACCESION	misc_feature
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	2241..3215 /gene="pbcC2" /locus_tag="Rv0929" 2241..3215 /gene="pbcC2" /locus_tag="Rv0929" /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT." /note="Rv0929, (MTCY21C12.23), len: 324 aa. pbcC2, phosphate-transport integral membrane ABC transporter (see citations below), highly similar to others e.g. NP_302394.1 NC_002677 membrane-bound component of phosphate transport from Mycobacterium leprae (319 aa); CAB86474.1 AF353816 phosphate ABC transport system permease protein from Streptomyces coelicolor (336 aa); NP_290359.1 NC_002655 high-affinity phosphate-specific transport system (cytoplasmic membrane component) from Escherichia coli strain 0157:H7 (319 aa); etc. Also similar to Rv935 MTCY08D9.04 PSTC1 PROBABLY TRANSMEMBRANE ABC TRANSPORTER COMPONENT OF PHOSPHATE UPTAKE SYSTEM from Mycobacterium tuberculosis (338 aa). Contains binding-protein-dependent transport systems inner membrane component signature (P500402)."
LOCUS	DEFINITION	ACCESION	gene
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	2241..3215 /gene="pbcC2" /locus_tag="Rv0929" 2241..3215 /gene="pbcC2" /locus_tag="Rv0929" /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT." /note="Rv0929, (MTCY21C12.23), len: 324 aa. pbcC2, phosphate-transport integral membrane ABC transporter (see citations below), highly similar to others e.g. NP_302394.1 NC_002677 membrane-bound component of phosphate transport from Mycobacterium leprae (319 aa); CAB86474.1 AF353816 phosphate ABC transport system permease protein from Streptomyces coelicolor (336 aa); NP_290359.1 NC_002655 high-affinity phosphate-specific transport system (cytoplasmic membrane component) from Escherichia coli strain 0157:H7 (319 aa); etc. Also similar to Rv935 MTCY08D9.04 PSTC1 PROBABLY TRANSMEMBRANE ABC TRANSPORTER COMPONENT OF PHOSPHATE UPTAKE SYSTEM from Mycobacterium tuberculosis (338 aa). Contains binding-protein-dependent transport systems inner membrane component signature (P500402)."
LOCUS	DEFINITION	ACCESION	misc_feature
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	2844..2930 /gene="pbcC2" /locus_tag="Rv0929" /note="P500402 Binding-protein-dependent transport systems inner membrane comp signature"
LOCUS	DEFINITION	ACCESION	gene
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	3212..4138 /gene="pbcA1" /locus_tag="Rv0930" 3212..4138 /gene="pbcA1" /locus_tag="Rv0930" /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE."
LOCUS	DEFINITION	ACCESION	misc_feature
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	2241..3215 /gene="pbcC2" /locus_tag="Rv0929" 2241..3215 /gene="pbcC2" /locus_tag="Rv0929" /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT." /note="Rv0929, (MTCY21C12.23), len: 324 aa. pbcC2, phosphate-transport integral membrane ABC transporter (see citations below), highly similar to others e.g. NP_302394.1 NC_002677 membrane-bound component of phosphate transport from Mycobacterium leprae (319 aa); CAB86474.1 AF353816 phosphate ABC transport system permease protein from Streptomyces coelicolor (336 aa); NP_290359.1 NC_002655 high-affinity phosphate-specific transport system (cytoplasmic membrane component) from Escherichia coli strain 0157:H7 (319 aa); etc. Also similar to Rv935 MTCY08D9.04 PSTC1 PROBABLY TRANSMEMBRANE ABC TRANSPORTER COMPONENT OF PHOSPHATE UPTAKE SYSTEM from Mycobacterium tuberculosis (338 aa). Contains binding-protein-dependent transport systems inner membrane component signature (P500402)."
LOCUS	DEFINITION	ACCESION	gene
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	2241..3215 /gene="pbcC2" /locus_tag="Rv0929" 2241..3215 /gene="pbcC2" /locus_tag="Rv0929" /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT." /note="Rv0929, (MTCY21C12.23), len: 324 aa. pbcC2, phosphate-transport integral membrane ABC transporter (see citations below), highly similar to others e.g. NP_302394.1 NC_002677 membrane-bound component of phosphate transport from Mycobacterium leprae (319 aa); CAB86474.1 AF353816 phosphate ABC transport system permease protein from Streptomyces coelicolor (336 aa); NP_290359.1 NC_002655 high-affinity phosphate-specific transport system (cytoplasmic membrane component) from Escherichia coli strain 0157:H7 (319 aa); etc. Also similar to Rv935 MTCY08D9.04 PSTC1 PROBABLY TRANSMEMBRANE ABC TRANSPORTER COMPONENT OF PHOSPHATE UPTAKE SYSTEM from Mycobacterium tuberculosis (338 aa). Contains binding-protein-dependent transport systems inner membrane component signature (P500402)."
LOCUS	DEFINITION	ACCESION	misc_feature
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	2844..2930 /gene="pbcC2" /locus_tag="Rv0929" /note="P500402 Binding-protein-dependent transport systems inner membrane comp signature"
LOCUS	DEFINITION	ACCESION	gene
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	3212..4138 /gene="pbcA1" /locus_tag="Rv0930" 3212..4138 /gene="pbcA1" /locus_tag="Rv0930" /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE."
LOCUS	DEFINITION	ACCESION	misc_feature
AP004112	Oryza sativa (japonica cultivar-group) chromosome 2 clone	AP004112	2241..3215 /gene="pbcC2" /locus_tag="Rv0929" 2241..3215 /gene="pbcC2" /locus_tag="Rv0929" /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT." /note="Rv0929

VERSION	AP004112.1	GI:15375098
KEYWORDS	HTG, HTGS_PHASE2.	
SOURCE	Oryza sativa (japonica cultivar-group)	
ORGANISM	Oryza sativa (japonica cultivar-group)	
	Euryzota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;	
	Spermatophytes, Magnoliophyta, Liliopsida, Poales, Poaceae;	
	Ehhartridoideae, Oryzaceae, Oryza.	
REFERENCE	1	
AUTHORS	Saasaki,T., Matsumoto,T. and Yamamoto,K.	
TITLE	Oryza sativa nippondare (GAJ) genomic DNA, chromosome 2, BAC clone:OJ1108_H04	
JOURNAL	Published Only in Database (2001)	
REFERENCES	2	(bases 1 to 144190)
AUTHORS	Saasaki,T., Matsumoto,T. and Yamamoto,K.	
TITLE	Direct Submission	
JOURNAL	Submitted (23-AUG-2001) Takuji Saasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan	
	(E-mail:tsaasaki@nias.affrc.go.jp, URL:http://xgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)	
COMMENT	The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.	
	NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.	
	* NOTE: This is a 'working draft' sequence.	
	* This sequence will be replaced	
	* by the finished sequence as soon as it is available and	
	* the accession number will be preserved.	
FEATURES	Location/Qualifiers	
SOURCE	1..144190	
	/organism="Oryza sativa (japonica cultivar-group)"	
	/mol_type="genomic DNA"	
	/cultivar="Nippondare"	
	/db_xref="taxon:39947"	
	/chromosome="2"	
	/clone="OJ1108_H04"	
ORIGIN		
	Query Match	84.8%; Score 17.8; DB 2; Length 144190;
	Best Local Similarity	90.5%; Pred. No. 2.9e+02;
	Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
Cy	1 TCCGATGATGATTTACCTCC 21	
Db	73873 TCCGATGAAGATTACCAACC 73893	
RESULT 9		
AP005649	166453 bp	linear
LOCUS	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,	
DEFINITION	BAC clone:OSUNBa0064G16.	
ACCESSION	AP005649	
VERSION	AP005649.3	GI:45736185
KEYWORDS		
SOURCE		
ORGANISM		
	Oryza sativa (japonica cultivar-group)	
	Oryza sativa (japonica cultivar-group)	
	Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;	
	Spermatophytes, Magnoliophyta, Liliopsida, Poales, Poaceae;	
	Ehhartridoideae, Oryzaceae, Oryza.	
REFERENCE	1	
AUTHORS	Saasaki,T., Matsumoto,T. and Katayose,Y.	
TITLE	Oryza sativa nippondare (GAJ) genomic DNA, chromosome 2, BAC clone:OSUNBa0064G16	
JOURNAL	Published Only in Database (2002)	
REFERENCES	2	(bases 1 to 166453)
AUTHORS	Saasaki,T., Matsumoto,T. and Katayose,Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-AUG-2002) Takuji Saasaki, National Institute of	

Agrobiological Sciences, Rice Genome Research Program; Kamonada 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail: saesaku@nias.affrc.go.jp, URL: <http://rsgp.dna.affrc.go.jp/>, Tel:81-298-38-7441, Fax:81-298-38-7468)  
On Mar 25, 2004 this sequence version replaced gi:2240965.  
Genes were predicted from the integrated results of the following:  
GENSCAN (<http://CCR-081.mtc.edu/GENSCAN.html>), GENESH (<http://www.softberry.com/>), GeneMark.hmm (<http://opal.biology.gatech.edu/GeneMark/>), glimmer3 ([http://www.tigr.org/tdb/glimmer/glim3\\_form.html](http://www.tigr.org/tdb/glimmer/glim3_form.html)), RiceANN (<http://rsgp.dna.affrc.go.jp/RiceANN/>), SplicePredictor (<http://bioinformatics.lasate.edu/cgi-bin/ep.cgi>), sam4 (<http://globin.cse.psu.edu/html/docs/sim4.html>), gapp2 (<http://www.tigr.org/software/glimmer/>). BLASTN and BLASTX. The genomic sequence was searched against NCBI Nonredundant Protein database, nr (<ftp://ncbi.nlm.nih.gov/blast/db/>) and the cDNA sequence database at RGP or DBJ. Protein homologs of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no.  
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.  
The orientation of the sequence is from -21M3 to M3rev of the BAC clone. This sequence of OSUNBA0064G16 clone has an overlap with OSUNBA0018G11 (DBJ: AP005776) clone at 5' end and with O1008\_C03 (DBJ: AP005288) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rsgp.dna.affrc.go.jp/genomeseq.html>.

Location/Qualifiers

1..166453  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="2"  
/clone="OSUNBA0064G16"  
338..3368  
/gene="OSUNBA0064G16.1"  
join(338..475,2574..2681,2803..2943,3030..3203,3297..23368)  
/gene="OSUNBA0064G16.1"  
/note="start and end point are not identified"  
join(338..475,2574..2681,2803..2943,3030..3203,3297..3368)  
/gene="OSUNBA0064G16.1"  
/note="Predicted by GeneMark.hmm etc."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD13231.1"  
/db\_xref="gi:45736186"  
/translation="MSVPIIITGYTAIDIMIDVIFINILNLFIAVSTATDRDLR  
HHEKDIHPLPVVAETLYTPDEIRLPLVOMFKGSMIMGMSGTEDPPTL  
RLTYNIGRETAANLVLYLESFASRPVDVLRKSNKATPPYATIRMKQLEEM  
QNGHSGVQCDKLKLHLSMSHQVVVDVDDVLRKSGKSGQS"  
join(4319..4454,6173..6241,6337..6636,6750..6851,  
7006..7200,7298..7585,7863..8049,9699..9745,9903..  
10264)  
/gene="OSUNBA0064G16.2"  
join(4319..4454,6173..6241,6337..6636,6750..6851,  
7006..7200,7298..7585,7863..8049,9699..9745,9903..10264)  
/gene="OSUNBA0064G16.2"  
/note="supported by full-length cDNA (s): AK063330"  
join(6182..6241,6337..6636,6750..6851,7006..7200,  
7298..7585,7863..8049,9699..9745,9903..9995)

/gene="OSJNBa0064G16.2"  
/note="contains full-length cDNA (s) : AK063330"  
/codon\_start=1  
/product="putative sucrose-phosphatase"  
/protein\_id="BAD1323.1"  
/db\_xref="GI:45736187"  
/translation="MDKLNGSARLMTIVSDLHTVVDHDEENLSLRFGALMEYVCO  
DSLIVFSGRSPTLYMELRKREKPHLTPITITMSGTEITTGSEBNVPDQGLVANKK  
DRNLIVESTANVSEIKLQVSEQRPHKVSFTVMSQAEVTKISSEKLEKGLVTKIT  
YSGQDDLDVLPQAGKQGLAVYLKQLSSCGKPPNNTLACDGSNDALFESIPGVHVG  
MVNSAOELLQWYSENAKDNPKIILATERCAAGIIQAIIGHFKLGPVNSPRVDPEYV  
ENPKPDVAVKPYLYEKMRARVPEKSDSTQYKNTTHANGVIHHPAGLECSLHAS  
IDALGSCYGDGCKGKKYAWDRILVNSCGSGMLVRFMLBEGVWSCCLTSLALNA  
KPTPEGRVYTHIKTKYKYSADBOGSKL"  
complement (8232. .8879)  
/gene="OSJNBa0064G16.3"  
complement (8292. .8879)  
/gene="OSJNBa0064G16.3"  
/note="probably inactive due to 5' exon missing in CDS  
pseudogene, callose synthase"  
pseudogene  
13379. .20883  
/gene="OSJNBa0064G16.4"  
join(13379. .13733,14288. .14362,14781. .14896,15076. .15190,  
15281. .15409,15730. .15777,16290. .16471,16549. .16745,  
16837. .17123,17310. .17459,17554. .18015,18122. .18271,  
18533. .18608,18772. .18899,19165. .19356,19480. .19611,  
19688. .19810,20049. .20267,20439. .20883)  
/gene="OSJNBa0064G16.4"  
/note="supported by full-length cDNA (s) : AK070600"  
join(13569. .13733,14288. .14362,14781. .14896,15076. .15190,  
15281. .15409,15730. .15777,16290. .16471,16549. .16745,  
16837. .17123,17310. .17459,17554. .18015,18122. .18271,  
18533. .18608,18772. .18899,19165. .19356,19480. .19611,  
19688. .19810,20049. .20267,20439. .20600)  
/gene="OSJNBa0064G16.4"  
/note="contains full-length cDNA (s) : AK070600"  
/codon\_start=1  
/product="cyclin-like protein"  
/protein\_id="BAD1323.1"  
/db\_xref="GI:45736188"  
/translation="MGSVGLNDITVSGVSRKVLPAACGLCYCPGLRARSPQYRX  
KIADIPATODEPERRRIKCEYVARNRHRPKITAYIEQCYKELRBRQYQPK  
VVLLIRKLVSCKQKPLASLSLITCTLDTRDMKICETLFDTVSQVVG  
TYQNLBELVPKCELAQIVAEKDKMLRASTQALASAMWIEGFSHISADENVY  
QVLESYVLQMQNDINDISEAPGRWYQVAKAGNATISRIPEKSLVDDGELHP  
AEDADENFMRVVCVHNAKLSREATRYVLSLFRFDNNNSMONTLAFCLLD  
MOIMENOGNIDIMISILVKEHKSVLKQPEMQLSVFVYIASLAEGRASAAAT  
GAISDLIRHMKTLHVALGSRDLVITKMDLGRANVDECTIQLSKRYGDAPVDDMS  
VMENISRTPLVATTAITSVTRIAQITISIPNLSIRNVFPEALPHQLLMVPHDH  
TRVSAHRIFSVVLVSSVSPFSKSTSPQLVHDIKTLSSAVSFSSALFPKLKR  
DKESFREKPDGSMNRSLSHADNDTSYKMDSSRSRSRSHKVPNFSKRVASLSLS  
PMSEKREKONTSAESCESTETLLRLSSROATLLSLIWAQISIPNTQNTYLAHTY  
SILLFSGSKATIFEALPSFOVAFSLMSVLEGTDLIPRRSLFLFLAMMFS  
RARNVAPLPIKCKMANERTMDPLHLVQDKLQVADCSERTYSGSPEDNNALKS  
AVELTQOSRESASTNNNTIRDLPSDLQTRQQLSDSFEDMCPISALFFELTVR  
NPGQDESSNQEDVLINMADNDTFGEVYENTATTSVSPANLIGIDELLESVTLTA  
PSQTRCSVSTAPNIPKEMTNOCEVLSMEKQCGSVSLSKHKGOSVGLINDADT  
GAHSSDDQNTPLQSLDGYPKYVADGALQVADVDVQQOFLKLPAASPDTFL  
KAAGC"  
22576. .22938  
/gene="OSJNBa0064G16.5"  
22576. .22938  
/gene="OSJNBa0064G16.5"  
/note="start and end point are not identified"  
22576. .22938  
/gene="OSJNBa0064G16.5"  
/note="predicted by RGENESH etc."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD1323.1"  
/db\_xref="GI:45736189"

/translation="MGYFRAPRRYAAAAGSESVALLVYGGGSEEAAPKGVAVVY  
GAERKRVFPTSTYRQPAFRGLMELADDFGFAOEGGILRLPCREEDFOATVADARR  
RPAAGGAIWSTMVARSU"  
complement (24680. .25238)  
/gene="OSJNBa0064G16.6"  
complement (join(24680. .24761,24856. .25238))  
/gene="OSJNBa0064G16.6"  
/note="start and end point are not identified"  
complement (join(24680. .24761,24856. .25238))  
/gene="OSJNBa0064G16.6"  
/codon\_start=1  
/product="auxin-induced protein-like"  
/protein\_id="BAD1323.1"  
/db\_xref="GI:45736190"  
/translation="MGSQGRASSNKIRIDIVRLHQLIKRMKRAALAPKAGNNNGGA  
SVPKGFPAVCVEEMERFVPTETLGMFAEQLRKAEERGFQHEALRIPCVEVF  
EGILRIVGRDOKAKMCSSESEHILTEEEFVGAELCYFQEDENVNHSF"  
complement (28164. .31674)  
/gene="OSJNBa0064G16.7"  
complement (join(28164. .29428,31041. .31674))  
/gene="OSJNBa0064G16.7"  
/note="start and end point are not identified"  
complement (join(28164. .29428,31041. .31674))  
/gene="OSJNBa0064G16.7"  
/note="similar to Oryza sativa chromosome 8,  
OJ1449.C01.27"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD1323.1"  
/db\_xref="GI:45736191"  
/translation="MRRSDKIVGATLVDPAHVAVLPSMWYKFKMTOSFAVGDLL  
VYMETGRALGCHPQRLRIIVKRGERVAKADMAPFRGRATVYEHGICRLSLAA  
DDGLVGMDSASTVADQPPPAQPCGVPFVPEFAPGTGNVHLGGGLCVAKLY  
MVAPGTCSPCCGGEEDMKFMALTGVEVARGGGELSIVKSKSCRYSGANGVNV

Query Match 84.8%; Score 17.8; DB 8; Length 166453;  
Best Local Similarity 90.5%; Pred. No. 36402; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 2;

Qy 1 TCCGATGATGATTACCTCC 21  
Db 114005 TCCGATGATGATTACCAACC 114025

RESULT 10  
AC079447 175567 bp DNA linear PRI 09-JAN-2002  
LOCUS  
DEFINITION Homo sapiens BAC clone Rpi1-111H13 from 2, complete sequence.  
AC079447  
AC079447.4 GI:15145541  
VERSION  
KEYWORDS  
HUG.  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 175567)  
Sulston,J.B. and Waterston,R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE  
99063792  
PUBMED  
9847074  
2 (bases 1 to 175567)  
Waterston,R.H.  
Submitted (01-SEP-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 175567)  
Waterston,R.H.

TITLE Direct Submission  
JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 175567)  
AUTHORS Waterston,R.  
JOURNAL Direct Submission  
Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Aug 9, 2001 this sequence version replaced g114550319.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: sapiens@wustl.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_NH011H13  
-----  
NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.  
  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.  
  
MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see http://genome.wustl.edu/gsc  
  
SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male  
donor, as described by Osceogawa,K., Moon,P.Y., Zhao,B., Frenken,B.,  
Tateno,M., Cataneese,J.V. and de Jong,P.V. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong  
and coworkers at the Roswell Park Cancer Institute  
(http://dscpac.med.buffalo.edu)  
VECTOR: pBAC3.6  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-38C17, the clone sequenced  
to the right is RP11-527J8, 2000 bp overlap. Actual start of this  
clone is at base position 1 of RP11-11H13; actual end is at base  
position 5330 of RP11-527J8.  
  
Data from AC023965 and AC018690 was used to finish this clone,  
AC079447.  
  
FEATURES  
source  
1. 175567  
location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-11H13"  
/clone\_1lb="RPCI-11"  
misc\_feature  
1. 1795  
/note="match to EST AL581351 (NID:g13948262) "  
misc\_feature  
1. 552  
/note="match to EST BG028419 (NID:g12417513) "  
misc\_feature  
1. 470  
/note="similar to Homo sapiens EST BG687463  
(NID:g1918860) "  
misc\_feature  
1. 463

misc\_feature  
1. 312  
/note="match to EST A1148460 (NID:g3676929) qc39e04.x1"  
repeat\_region  
3. 123  
/note="match to EST AM391243 (NID:g6896006) "  
misc\_feature  
15. 763  
/rpt\_family="L2"  
misc\_feature  
21. 457  
/note="match to EST BE620779 (NID:g9891717) "  
misc\_feature  
42. 792  
/note="match to EST AV703935 (NID:g10733284) "  
misc\_feature  
/note="match to EST BG742002 (NID:g14052655) "  
misc\_feature  
211. 970  
/note="match to EST BG284097 (NID:g13034702) "  
misc\_feature  
214. 393  
/note="match to EST W02815 (NID:g1274948) za07d03.r1"  
misc\_feature  
250. 817  
/note="match to EST A1309762 (NID:g4004633) qo75a08.x1"  
misc\_feature  
234. 732  
/note="match to EST AA459888 (NID:g2184795) zx65h10.r1"  
misc\_feature  
353. 711  
/note="match to EST AA459888 (NID:g2184795) zx65h10.r1"  
misc\_feature  
/note="similar to Homo sapiens EST R28040 (NID:g784175)  
yh6b03.r1"  
misc\_feature  
358. 930  
/note="match to EST W37953 (NID:g1319547) zc13e02.r1"  
misc\_feature  
419. 798  
/note="match to Homo sapiens EST R28262 (NID:g784397)  
yh6b03.r1"  
misc\_feature  
469. 845  
/note="match to EST T71834 (NID:g686355) yc62g12.r1"  
misc\_feature  
515. 785  
/note="match to EST AV703935 (NID:g10733284) "  
misc\_feature  
526. 998  
/note="match to EST AA633739 (NID:g2556953) ag87f03.r1"  
misc\_feature  
527. 796  
/note="match to EST BG214283 (NID:g137430304) "  
misc\_feature  
529. 1134  
/note="match to EST BE256300 (NID:g9126854) "  
misc\_feature  
863  
/note="match to EST T71834 (NID:g686355) yc62g12.r1"  
misc\_feature  
883. 884  
/note="similar to Homo sapiens EST R28262 (NID:g784397)  
yh6b03.r1"  
misc\_feature  
989. 1680  
/note="match to EST BE535937 (NID:g9764582) "  
repeat\_region  
1019. 1316  
/rpt\_family="Alu"  
misc\_feature  
1220. 1944  
/note="match to EST BG495526 (NID:g12457041) "  
repeat\_region  
1523. 2003  
/rpt\_family="MAAR"  
misc\_feature  
1617. 2399  
/note="match to EST BG613322 (NID:g1364693) "  
misc\_feature  
1729. 2117  
/note="match to EST AA773220 (NID:g2824791) af78e04.r1"  
misc\_feature  
1877. 2332  
/note="match to EST AW297722 (NID:g6704358) "  
misc\_feature  
2238. 2239  
/note="match to EST A1830655 (NID:g5451326) wj10d06.x1"  
misc\_feature  
2455. 2931  
/note="match to EST A1830655 (NID:g5451326) wj10d06.x1"  
misc\_feature  
2456. 2634  
/note="match to EST AW798659 (NID:g7850542) "  
misc\_feature  
2847. 3164  
/note="match to EST AA971709 (NID:g3146999) op95b11.r1"  
repeat\_region  
3043. 3184  
/rpt\_family="Alu"  
misc\_feature  
3099. 3330  
/note="match to EST BF811752 (NID:g12140940) "  
misc\_feature  
3525. 4128  
/note="match to EST AL135413 (NID:g6603600) "  
misc\_feature  
3632. 3633  
/note="match to EST BF811752 (NID:g12140940) "  
repeat\_region  
3747. 4023

```

repeat_region      /rpc_family="Alu"
                    4330..4485
repeat_region      /rpc_family="ERV1"
                    4528..4827
repeat_region      /rpc_family="ERV1"
                    4994..5511
repeat_region      /rpc_family="ERV1"
                    5891..5918
repeat_region      /rpc_family="AT-rich"
                    5905..6091
repeat_region      /rpc_family="L1"
                    6092..6352
repeat_region      /rpc_family="Alu"
                    6339..6372
repeat_region      /rpc_family="AT-rich"
                    6353..7447
repeat_region      /rpc_family="L1"
                    7439..7510
repeat_region      /rpc_family="CT-rich"
                    7489..7724
repeat_region      /rpc_family="Alu"
                    7727..8263
repeat_region      /rpc_family="L1"
                    8260..8295
repeat_region      /rpc_family="(T)n"
                    8266..8523
repeat_region      /rpc_family="Alu"

```

```

Query Match      84.8%; Score 17.8; DB 9; Length 175567;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 TCCGAGATGATTTACTCTCC 21
Db      86970 TCCGCGTGAATTAATTAATCTCC 86990

```

```

RESULT 11
AC092002/c      195744 bp DNA linear HTG 07-SEP-2001
LOCUS      Homo sapiens chromosome RP11-364D20, WORKING DRAFT
DEFINITION
AC092002
AC092002.2 GI:15487482
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS      Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Eukaryota; Euteleostomi; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 195744)
              Waterston,R.H.
              The sequence of Homo sapiens clone
              Unpublished
              2 (bases 1 to 195744)
              Waterston,R.H.
              Direct Submission
              Submitted (11-JUN-2001) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
              On Sep 7, 2001 this sequence version replaced gi:14336624.
COMMENT
              ----- Genome Center -----
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc/index.shtml
              ----- Project Information -----
              Center project name: H NH0364D20
              ----- Summary Statistics -----
              Sequencing vector: M13; 12%
              Sequencing vector: plasmid; 88%
              Chemistry: Dye-terminator Big Dye; 94% of reads

```

```

Assembly program: Phrap; version 0.990319
Consensus quality: 189590 bases at least Q40
Consensus quality: 191694 bases at least Q30
Consensus quality: 192828 bases at least Q20

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 11257: contig of 11257 bp in length
* 11258: gap of unknown length
* 11358: contig of 1616 bp in length
* 27973: gap of unknown length
* 28073: gap of unknown length
* 28074: contig of 21597 bp in length
* 49671: gap of unknown length
* 49770: gap of unknown length
* 49771: contig of 14598 bp in length
* 64369: gap of unknown length
* 64468: gap of unknown length
* 83993: contig of 19525 bp in length
* 83994: gap of unknown length
* 84093: gap of unknown length
* 84094: contig of 24530 bp in length
* 108623: gap of unknown length
* 108724: contig of 29220 bp in length
* 137943: gap of unknown length
* 137944: gap of unknown length
* 138044: contig of 46828 bp in length
* 184871: gap of unknown length
* 184972: gap of unknown length
* 184973: contig of 2065 bp in length
* 187037: gap of unknown length
* 187137: gap of unknown length
* 187137 195744: contig of 8608 bp in length.

```

## FEATURES

## Source

```

1..195744
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="RP11-364D20"
1..11257
/contig="RP11-364D20"
11258..27973
/contig="assembly_name:Contig10"
11358..27973
/contig="assembly_name:Contig11"
28074..49671
/contig="assembly_name:Contig12"
49771..64368
/contig="assembly_name:Contig13"
64469..83993
/contig="assembly_name:Contig14"
84094..108623
/contig="assembly_name:Contig15"
108724..137943
/contig="assembly_name:Contig16"
138044..184871
/contig="assembly_name:Contig17"
184972..187036
/contig="assembly_name:Contig18"
187137..195744
/contig="assembly_name:Contig19"

```

## ORIGIN

```

Qy      1 TCCGAGATGATTTACTCTCC 21
Db      16856 TCCGCGTGAATTAATTAATCTCC 16836
Query Match      84.8%; Score 17.8; DB 2; Length 195744;
Best Local Similarity 90.5%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

RESULT 12  
AC112285/c  
LOCUS  
DEFINITION  
AC112285 247746 bp DNA linear HTG 13-MAY-2003  
\*\*\* 2 unordered pieces.  
AC112285  
AC112285.4 GI:30580789  
HTG: HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eutheria; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 247746)  
Munry, D., Marle, M., Metzer, M., Lee, A., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Deigado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Driper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C., A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C., M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gargese, J., E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M.,  
Guarinate, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hognes, M.,  
Hollins, B., Howell, S., Hulix, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kovacs, C., Kraft, C., L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenz, H., L., Louie, H., Lozano, R., J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mamoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mathew, S., McLeod, M., P., McNeill, T., Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Mundana, M., Murphy, M., Nair, L.,  
Markovits, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwokwelen, O., Okunnu, G., Olampunsa, A., Pal, S., Parks, K.,  
Paetel, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Polidexter, A., Popovic, D., Primus, B., Pu, L.,  
Puzo, M., Qutroz, J., Rachlin, R., Reeves, K., Regier, M., A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rivers, C., Rodkey, T., Rojce, A., Rose, M., Rose, R., Ruiz, S., J.,  
Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
Shetty, J., Sivatsbeyn, A., Sisson, I., Sitter, C., D., Smajic, D.,  
Sned, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Tjofes, Z., Umami, K.,  
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wiczyk, R., Woodson, H., Worley, K.,  
Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, Y., Yoon, L., Yoon, V.,  
Niederhauser, A., Weiser, R., Smith, D., R., Holt, R., A., Smith, H., O.,  
Weinstock, G., and Gibbs, R. A.  
Direct Submission  
Unpublished  
2 (bases 1 to 247746)  
Worley, K. C.  
Direct Submission  
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 247746)

AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On May 13, 2003 this sequence version replaced gi:23664473.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: G1XQ  
Center clone name: CH230-97H21  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 236310 bases at least Q40  
Consensus quality: 236673 bases at least Q30  
Consensus quality: 240367 bases at least Q20  
Estimated insert size: 248967; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a working draft sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 246339; contig of 246339 bp in length  
\* 246340 247746; gap of unknown length  
\* 246440 247746; contig of 1307 bp in length.  
FEATURES  
source  
1. 247746  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-97H21"  
1. 11764  
/note="wgs\_end-extension  
/note="wgs\_end-5p6"  
6257..6913  
/note="clone boundary  
clone\_end:5p6  
site:ECORI  
end\_sequence:BH282614"  
150844..151647  
/note="clone boundary  
clone\_end:T7  
site:ECORI  
end\_sequence:BH282612"  
ORIGIN  
Query Match 84.8%; Score 17.8; DB 2; Length 247746;  
Best Local Similarity 90.5%; Pred.No. 3.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TCCGATGATGATTACTCC 21



DB 154435 TCAGATGATGTTTACTCC 154415

# RESULT 13 AC128496/C LOCUS

AC128496 291544 bp DNA linear HTG 19-NOV-2002  
Rattus norvegicus clone CH230-444D9, WORKING DRAFT SEQUENCE, 3  
unordered pieces.

AC128496.3 GI:25085139  
HTG, HTGS, PHASE1, HTGS DRAFT, HTGS FULLTOP.  
Rattus norvegicus (Norway rat)  
ORGANISM

# REFERENCE AUTHORS

1 (bases 1 to 291544)  
Murny,D,Marie,,Metzker,M,Lee,,Abramson,S,,Adams,C,,Alder,J,,  
Allen,C,Allen,H,,Albrooke,S,,Amin,A,,Angiano,D,,  
Anyalebechi,V,,Aoyagi,A,,Ayodeji,M,,Baca,E,,Baden,H,,  
Baldwin,D,,Bandaraike,D,,Barber,M,,Barnstead,M,,Benahmed,F,,  
Biswal,K,,Blair,J,,Blankenburg,K,,Blyth,P,,Brown,M,,  
Bryant,N,,Buhay,C,,Burch,P,,Burrell,K,,Calderon,E,,  
Cardenas,V,,Carter,K,,Cavazos,I,,Cesari,H,,Center,A,,  
Chacko,J,,Chavez,D,,Chen,G,,Chen,R,,Chen,Y,,Chen,Z,,Chu,J,,  
Cleveland,C,,Cockrell,R,,Cox,C,,Coyle,M,,Cree,A,,D'Souza,L,,  
Daviila,M,L,,Davis,C,,Davy-Carroll,U,,De Ande,C,,Dedrich,D,,  
Delgado,O,,Denson,S,,Deramo,C,,Ding,Y,,Dinh,H,,Divya,K,,  
Draper,H,,Dugan-Rocha,S,,Dunn,A,,Durbin,K,,Duval,B,,Eaves,K,,  
Egan,A,,Escotto,M,,Eugene,C,,Evans,C,A,,Falls,T,,Fan,G,,  
Fernandez,S,,Finley,M,,Flagg,N,,Forbes,L,,Foster,M,,Foster,P,,  
Fraser,C,M,,Gabisi,A,,Ganta,R,,Garcia,A,,Garner,T,,Garram,M,,  
Geregegoxis,B,,Geer,K,,Gill,R,,Grady,M,,Guerra,W,,Guevara,W,,  
Gunnaratne,P,,Haaland,W,,Hamil,C,,Hamilton,C,,Hamilton,K,,  
Harvey,Y,,Havlik,P,,Hawes,A,,Henderson,N,,Hernandez,J,,  
Hernandez,R,,Hines,S,,Hladun,S,L,,Hodgson,A,,Hogues,M,,  
Hollins,B,,Howells,S,,Hulky,S,,Hume,J,,Idlebirt,D,,Jackson,A,,  
Jackson,L,,Jacob,L,,Jiang,H,,Johnson,B,,Johnson,R,,Jollivet,A,,  
Karpathy,S,,Kelly,S,,Kelly,S,,Khan,Z,,King,L,,Kovar,C,,  
Kovats,C,,Kraft,C,L,,Lebow,H,,Levan,J,,Lewis,L,,Li,Z,,Liu,J,,  
Liu,J,,Liu,W,,Liu,Y,,London,P,,Longacre,S,,Lopez,J,,  
Lorenshewa,L,,Louisege,H,,Lozando,R,J,,Lu,X,,Ma,J,,  
Maheshwari,M,,Mahindartne,M,,Mahmoud,M,,Malloy,K,,Mangum,A,,  
Mangum,B,,Mapua,P,,Martin,K,,Martin,R,,Martinez,E,,  
Mathew,J,S,,McLeod,M,P,,McNeill,T,Z,,Meenen,B,,  
Milosavljevic,A,,Miner,G,,Minja,E,,Montemayor,J,,Moore,S,,  
Morgan,M,,Morris,K,,Morris,S,,Mundasa,M,,Murphy,M,,Nair,L,,  
Narkervis,C,,Neal,D,,Newton,N,,Nguyen,N,,Norris,S,,  
Nwokilembh,O,,Okunnu,G,,Olarunpasegun,A,,Pal,S,,Parke,K,,  
Pacternak,S,,Paul,H,,Perez,A,,Perez,L,,Plamkoch,C,,  
Plopper,F,,Polndexter,A,,Popovic,D,,Primus,E,,Pu,L,-,,  
Puzo,M,,Quiroz,J,,Rachlin,E,,Reeves,K,,Rejzler,M,A,,Reigh,R,,  
Reilly,B,,Reilly,M,,Ren,Y,,Reuter,M,,Richards,S,,Riggs,F,,  
Rivers,C,,Rockey,T,,Rojas,A,,Rose,M,,Rose,R,,Ruiz,S,J,,  
Sanders,W,,Savery,G,,Scherer,S,,Scott,G,,Shatman,S,,Shen,H,,  
Shetty,J,,Shvartsbeyn,A,,Slason,I,,Slitter,C,D,,Smajic,D,,  
Sneed,A,,Sodergren,E,,Song,X,-Z,,Sorelle,R,,Sosa,J,,  
Steimle,M,,Strong,R,,Sutlon,A,,Svatek,A,,Taboi,P,,Taylor,C,,  
Taylor,T,,Thomas,N,,Thomas,S,,Tingey,A,,Treylos,Z,,Uemami,K,,  
Valas,R,,Vera,V,,Villasana,D,,Waldron,L,,Walker,B,,Wang,J,,  
Wang,O,,Wang,S,,Warren,J,,Warren,R,,Wei,X,,White,F,,  
Williams,G,,Willson,R,,Wiczek,R,,Wooden,H,,Worley,K,,  
Wright,D,,Wright,R,,Wu,J,,Yakub,S,,Yen,J,,Yoon,L,,Yoon,V,,  
Yu,F,,Zhang,J,,Zhou,J,,Zhou,X,,Zhao,S,,Zhu,D,,von  
Niederhausern,A,,Weiss,R,,Smith,D,R,,Holt,R,A,,Smith,H,O,,  
Weinstock,G, and Gibbs,R.A.  
Unpublished  
Direct Submission  
2 (bases 1 to 291544)  
Worley,K.C.  
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department

# REFERENCE AUTHORS TITLE JOURNAL

# COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 291544)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23907748.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

# ----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GYSW

Center clone name: CH230-444D9

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 187201 bases at least Q40

Consensus quality: 189849 bases at least Q30

Consensus quality: 191486 bases at least Q20

Estimated insert size: 193264; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 288506: contig of 288506 bp in length

\* 288507 288606: gap of unknown length

\* 288607 289949: contig of 1343 bp in length

\* 289950 290049: gap of unknown length

\* 290050 291544: contig of 1495 bp in length.

Location/Qualifiers

1..291544

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-444D9"

1..1841

/note="wgs end extension

/clone\_end:"7"

6777..7677

/note="clone boundary

/clone\_end:"7"

site:

end\_sequence:"BZ198832"

complement(187560..188645)

/note="clone boundary

clone\_end:Sp6

end\_sequence:"BZ198833"

189481..192835

/note="wgs\_end extension

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature



misc\_feature clone\_end:Sp6"  
195005..196648  
/note="wgs\_end\_extension  
clone\_end:Sp6"

ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 291544;  
Best Local Similarity 90.5%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTACTCC 21  
|||||  
Db 10095 TCAGATGATGTTACTCC 10075  
|||||

RESULT 14  
AC149573 320731 bp DNA linear HTG 06-OCT-2004  
LOCUS Medicago truncatula clone mth2-119g4, WORKING DRAFT SEQUENCE, 52  
DEFINITION unordered pieces.  
AC149573  
AC149573.8 GI:53828780  
HTG; HTGS PHASE1; HTGS DRAFT.  
KEYWORDS Medicago truncatula (barrel medic)  
SOURCE Medicago truncatula  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;  
Medicago.  
1 (bases 1 to 320731)  
Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.  
and Roe,B.A.  
2 Medicago truncatula BAC Clone mth2-119g4  
Unpublished  
2 (bases 1 to 320731)  
Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.  
and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (08-JUN-2004) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
3 (bases 1 to 320731)  
Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.  
and Roe,B.A.  
REFERENCE Submitted (06-OCT-2004) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
COMMENT On Oct 6, 2004 this sequence version replaced gi:53793764.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 52 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2120: contig of 2120 bp in length  
\* 2121 2220: gap of unknown length  
\* 2221 4377: contig of 2157 bp in length  
\* 4378 4477: gap of unknown length  
\* 4478 6499: contig of 2022 bp in length  
\* 6499 6599: gap of unknown length  
\* 6599 8658: contig of 2059 bp in length  
\* 8658 8758: gap of unknown length  
\* 8758 11276: contig of 2518 bp in length  
\* 11276 11376: gap of unknown length  
\* 11376 13990: contig of 2614 bp in length  
\* 13990 14090: gap of unknown length  
\* 14090 16332: contig of 2242 bp in length  
\* 16332 16432: gap of unknown length  
\* 16432 18500: contig of 2068 bp in length  
\* 18500 18600: gap of unknown length  
\* 18600 21151: contig of 2551 bp in length  
\* 21151 21251: gap of unknown length  
\* 21251 23489: contig of 2238 bp in length  
\* 23489 23589: gap of unknown length  
\* 23589 26707: contig of 3118 bp in length  
\* 26707 26807: gap of unknown length  
\* 26807 30025: contig of 3218 bp in length  
\* 30025 30125: gap of unknown length  
\* 30125 33133: contig of 3008 bp in length  
\* 33133 33233: gap of unknown length  
\* 33233 35396: contig of 2163 bp in length  
\* 35396 35496: gap of unknown length  
\* 35496 38285: contig of 2789 bp in length  
\* 38285 42217: gap of unknown length  
\* 42217 42317: contig of 3832 bp in length  
\* 42317 44932: gap of unknown length  
\* 44932 45032: contig of 2615 bp in length  
\* 45032 48055: gap of unknown length  
\* 48055 48155: gap of unknown length  
\* 48155 50571: contig of 2416 bp in length  
\* 50571 50671: gap of unknown length  
\* 50671 52786: contig of 2115 bp in length  
\* 52786 52886: gap of unknown length  
\* 52886 55014: contig of 2128 bp in length  
\* 55014 55114: gap of unknown length  
\* 55114 58456: contig of 3342 bp in length  
\* 58456 58556: gap of unknown length  
\* 58556 60674: contig of 2018 bp in length  
\* 60674 60735: gap of unknown length  
\* 60735 60933: contig of 4259 bp in length  
\* 60933 65033: gap of unknown length  
\* 65033 66822: contig of 3589 bp in length  
\* 66822 68723: gap of unknown length  
\* 68723 73490: contig of 4768 bp in length  
\* 73490 73590: gap of unknown length  
\* 73590 77278: contig of 3688 bp in length  
\* 77278 80719: gap of unknown length  
\* 80719 80819: contig of 3341 bp in length  
\* 80819 85277: gap of unknown length  
\* 85277 85378: contig of 4458 bp in length  
\* 85378 92307: gap of unknown length  
\* 92307 92406: contig of 6929 bp in length  
\* 92406 97236: gap of unknown length  
\* 97236 97337: contig of 4830 bp in length  
\* 97337 102561: gap of unknown length  
\* 102561 102661: contig of 5225 bp in length  
\* 102661 106887: gap of unknown length  
\* 106887 106988: contig of 4226 bp in length  
\* 106988 112013: gap of unknown length  
\* 112013 112113: contig of 5026 bp in length  
\* 112113 115962: gap of unknown length  
\* 115962 117062: contig of 4849 bp in length  
\* 117062 122443: gap of unknown length  
\* 122443 122443: contig of 5181 bp in length  
\* 122443 131655: gap of unknown length  
\* 131655 131755: contig of 9312 bp in length  
\* 131755 137837: gap of unknown length  
\* 137837 137937: contig of 6082 bp in length  
\* 137937 144111: gap of unknown length  
\* 144111 144211: contig of 6174 bp in length  
\* 144211 153611: gap of unknown length  
\* 153611 153611: contig of 9300 bp in length  
\* 153611 159483: gap of unknown length  
\* 159483 159583: contig of 5872 bp in length  
\* 159583 168698: gap of unknown length  
\* 168698 168798: contig of 9115 bp in length  
\* 168798 168998: gap of unknown length

```
* 168799 178460: contig of 9662 bp in length
* 178461 178560: gap of unknown length
* 178561 189146: contig of 10586 bp in length
* 189147 189246: gap of unknown length
* 189247 205996: contig of 16750 bp in length
* 205997 206097: gap of unknown length
* 206097 221512: contig of 15416 bp in length
* 221513 221612: gap of unknown length
* 221613 235904: contig of 14292 bp in length
* 235905 236005: gap of unknown length
* 236005 248142: contig of 12138 bp in length
* 248143 248242: gap of unknown length
* 248243 263641: contig of 15399 bp in length
* 263642 263741: gap of unknown length
* 263742 280695: contig of 16954 bp in length
* 280696 280795: gap of unknown length
* 280796 301898: contig of 21103 bp in length
* 301899 301998: gap of unknown length
* 301999 320731: contig of 18733 bp in length.
```

## FEATURES

source

```
1..320731
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-119g4"
/clone_lib="Medicago truncatula BAC library H2"
```

## ORIGIN

```
Query Match      84.8%; Score 17.8; DB 2; Length 320731;
Best Local Similarity 90.5%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      1 TCCGGATGATGATTACCTC 21
Db      95671 TCCGGATGACGAATTCCTCC 95691
```

```
RESULT 15
ATTU90428/c      1286 bp DNA linear PLN 25-MAR-1997
LOCUS      Arabidopsis thaliana blue copper-binding protein II gene, complete
DEFINITION cda.
```

ACCESSION U90428 GI:1905999

## KEYWORDS

Arabidopsis thaliana (chale cress)

Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 1286)

Kim,C.H., Cho,Y.H. and Hong,Y.-N.

Direct Submission Submitted (22-FEB-1997) Department of Biology, College of Natural Science, Seoul National University, 56-1 Shillim-Dong, Kwanak-Ku, Seoul 151-742, Korea

location/Qualifiers

## FEATURES

source

```
1..1286
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/joint(160..331,774..1207)
/note="BCB II"
/codon_start=1
/product="blue copper-binding protein II"
/protein_id="AAB50232.1"
/db_xref="GI:1906000"
/translation="MAWNGLSKMAVAAATALLVLTTPGAVAVTYTTEMTGVDSG
WATKTRPVGILIEFKYGSHTVVDKAGDGDGASSTENHSDGDTKIDLKTVGIN
YFICSTPGHCSLNGMKLAVNVVAGSADLRTPSPSPGTPPTPESSPGSGSPPTT
PTPGASTSPPPKASGASGXVMSYLVGVSMVLGYGLMM"
```

## CDS

## ORIGIN

***This Page Blank (uspio)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21 ; Search time 242.17 Seconds  
(without alignments)  
513.336 Million cell updates/sec

Title: US-10-777-131a-1

Perfect score: 21

Sequence: 1 tccgagatgattactctcc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	8	ABZ79927 Mycobacte
2	21	100.0	823	8	ABZ79951 Mycobacte
3	21	100.0	3002	13	ADZ57503 Mycobacte
4	21	100.0	110000	4	AA199682_12 Continuation (13 o
5	21	100.0	110000	4	AA199683_12 Continuation (13 o
6	17.4	82.9	1275	8	ACA47360 Prokaryot
7	17	81.0	397	3	AAA15451 Aaals451 DNA seque
8	17	81.0	110000	10	ADF77343_03 Continuation (4 of
9	16.8	80.0	131239	12	ADQ21179 Human sof
10	16.2	77.1	225	9	AC192883 Human mlc
11	16.2	77.1	288	9	AC11729 Arabidops
12	16.2	77.1	304	9	ADA58045 Maize trl
13	16.2	77.1	304	12	ADP60670 Maize car
14	16.2	77.1	316	9	ADA57946 Maize trl
15	16.2	77.1	316	12	ADP60571 Maize car
16	16.2	77.1	325	12	ADP60572 Maize car
17	16.2	77.1	325	12	ADP60572 Maize car
18	16.2	77.1	398	9	ADA58043 Maize trl
19	16.2	77.1	398	12	ADP60668 Maize car
20	16.2	77.1	466	3	AA344481 Arabidops

21	16.2	77.1	1116	8	ACA20952 Prokaryot
22	16.2	77.1	1131	3	ADA30615 DNA encod
23	16.2	77.1	1221	3	AA343666 Zea mays
24	16.2	77.1	1329	8	ACA30420 Prokaryot
25	16.2	77.1	3435	6	ABZ66862 Arabidops
26	16.2	77.1	4647	6	ABA90426 Drosophil
27	16.2	77.1	5316	4	AB111843 Drosophil
28	16.2	77.1	7499	4	AB111842 Drosophil
29	16.2	77.1	14775	4	AB117452 Drosophil
30	16.2	77.1	110000	12	ADN46845_06 Continuation (7 of
31	16.2	77.1	110000	12	ADN47591_14 Continuation (7 of
32	16.2	77.1	110000	12	ADN46123_06 Continuation (7 of
33	16.2	77.1	110000	12	ADN47209_14 Continuation (7 of
34	16.2	77.1	110000	12	ADN46464_06 Continuation (7 of
35	16.2	77.1	110000	12	ADN47960_14 Continuation (15 o
36	16	76.2	1080	8	ACA31076 Prokaryot
37	16	76.2	3510	3	AA656299 Maize sca
38	15.8	75.2	434	12	ADQ20901 Human sof
39	15.8	75.2	557	12	ADQ24851 Human sof
40	15.8	75.2	1215	13	ADT41714 Bacterial
41	15.8	75.2	4398	6	ADH32179 Yeast sma
42	15.8	75.2	110000	3	AA622305_00 Arabidops
43	15.8	75.2	110000	3	AA622305_01 Continuation (2 of
44	15.8	75.2	116624	2	AAV52850 Human eya
45	15.4	73.3	216	2	AA12103 Human bia

## ALIGNMENTS

RESULT 1	ABZ79927	standard; DNA; 21 BP.
ID	ABZ79927	
AC	ABZ79927	
XX	19-MAY-2003 (first entry)	
DT		
XX		
DE	Mycobacterium tuberculosis mutT2-1 PCR primer SEQ ID NO:1.	
XX		
KW	Mycobacterium tuberculosis; mutT2; alkA; ogt; RV3908; mutY; RV3909;	
KW	detection; multidrug resistance; multiple drug resistance; MDR;	
XX	infection; PCR primer; ss.	
OS	Mycobacterium tuberculosis.	
OS	Synthetic.	
PN	WO2003016562-A2.	
XX		
PD	27-FEB-2003.	
XX		
PF	14-AUG-2002; 2002WO-EP009679.	
XX		
PR	14-AUG-2001; 2001US-0311824P.	
PR	21-AUG-2001; 2001US-0313523P.	
XX		
PA	(INSP ) INST PASTEUR.	
XX		
PI	Gicquel B;	
XX		
DR	WPI, 2003-256711/25.	
XX		
PT	Predicting the epidemic character of a Mycobacterium tuberculosis isolate	
PT	and/or the acquisition of multiple drug resistance (MDR) by the isolate	
PT	by detecting an alteration in the DNA repair system of the isolate.	
XX		
PS	Claim 32; Page 16; 83pp; English.	
XX		
CC	The present invention describes a method for predicting the epidemic	
CC	character of a Mycobacterium tuberculosis isolate and/or a selective	
CC	advantage to be maintained in the host and/or the acquisition of multiple	
CC	drug resistance (MDR) by the isolate comprising detecting an alteration	
CC	in the DNA repair system of the isolate. Also described: (1) detecting a	

CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a PCR primer for M. tuberculosis mutT2,  
CC which is used in the exemplification of the present invention  
XX  
SQ Sequence 21 BP; 4 A; 6 C; 4 G; 7 T; 0 U; 0 Other;  
Query Match 100.0%; Score 21; DB 8; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.81;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCCGGATGATGATTACTCTCC 21  
DB 1 TCCGGATGATGATTACTCTCC 21  
RESULT 2  
ABZ79951  
ID ABZ79951 standard; DNA; 823 BP.  
XX  
AC ABZ79951;  
XX  
DT 19-MAY-2003 (first entry)  
XX  
DE Mycobacterium tuberculosis mutT2 nucleotide sequence SEQ ID NO.19.  
XX  
XX Mycobacterium tuberculosis; mutT2; alka; ogt; RV3908; mutY; RV3909;  
KW detection; multidrug resistance; multiple drug resistance; MDR;  
KW infection; gene; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO2003016562-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002MO-EP009679.  
XX  
PR 14-AUG-2001; 2001US-0311824P.  
XX  
PR 21-AUG-2001; 2001US-0313523P.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Glcquel B;  
XX  
DR WPI, 2003-256711/25.  
XX  
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate  
PT by detecting an alteration in the DNA repair system of the isolate.  
XX  
PS Disclosure; Fig 4A; 83pp; English.  
XX  
CC The present invention describes a method for predicting the epidemic  
CC character of a Mycobacterium tuberculosis isolate and/or a selective  
CC advantage to be maintained in the host and/or the acquisition of multiple  
CC drug resistance (MDR) by the isolate comprising detecting an alteration  
CC in the DNA repair system of the isolate. Also described: (1) detecting a  
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a M. tuberculosis mutT2 nucleotide

CC sequence, which is used in the exemplification of the present invention  
XX  
SQ Sequence 823 BP; 144 A; 253 C; 277 G; 149 T; 0 U; 0 Other;  
Query Match 100.0%; Score 21; DB 8; Length 823;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCCGGATGATGATTACTCTCC 21  
DB 5 TCCGGATGATGATTACTCTCC 25  
RESULT 3  
ADS75303  
ID ADS75303 standard; DNA; 3002 BP.  
XX  
AC ADS75303;  
XX  
DT 16-DEC-2004 (first entry)  
XX  
DE Mycobacterium tuberculosis nargHJI nitrate reductase operon Segid 1.  
XX  
XX nargHJI nitrate reductase operon; single nucleotide polymorphism; SNP;  
KW detection method; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
FH Key Location/Qualifiers  
FT variation replace(1500,C)  
FT /\*tag= a  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT mlec\_feature 1717..1719  
FT /\*tag= b  
FT /\*note= "GTC start codon"  
XX  
PN WO2004083459-A1.  
XX  
PD 30-SEP-2004.  
XX  
PF 19-MAR-2004; 2004MO-EP002911.  
XX  
PR 20-MAR-2003; 2003DE-01013791.  
XX  
PA (ARTU-) ARTUS GBS MOLEKULARBIOLOGISCHE DIAGNOSTI.  
XX  
PI Bange F;  
XX  
DR WPI, 2004-728487/71.  
XX  
PT Specific detection of Mycobacterium tuberculosis, useful for differential  
PT diagnosis of infection, by detecting specific polymorphism in the  
PT promoter of the nitrate reductase operon.  
XX  
PS Claim 1; SEQ ID NO 1; 46pp; German.  
XX  
CC This invention relates to a novel method for the detection of  
CC Mycobacterium tuberculosis (M. tuberculosis) occurring in a biological  
CC sample. Specifically, it refers to the amplification of the nargHJI  
CC nitrate reductase operon, including position -215 in the 5' to 3' reading  
CC direction that occurs upstream of the GTC start codon of the operon. The  
CC present invention describes detecting an M. tuberculosis specific  
CC polymorphism at position -215. The polymorphism (C-215T) is detected by  
CC hybridisation to specific probes such that M. tuberculosis can be  
CC differentiated from M. bovis, M. bovis BCG, M. africanum and M. microti  
CC by analysis of the melting curve. The method is used to detect M.  
CC tuberculosis infection, especially to differentiate it from other species  
CC of the M. tuberculosis complex, and furthermore provides rapid and  
CC specific detection, and differentiation of M. tuberculosis. This  
CC polymnucleotide sequence is the M. tuberculosis nargHJI nitrate reductase  
CC operon containing the -215 SNP at position 1500 in this sequence of the  
CC invention.  
XX

SQ Sequence 3002 BP; 483 A; 923 C; 1013 G; 583 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 13; Length 3002;

Best Local Similarity 100.0%; Pred. No. 1; 7; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCCGATGATGATTTACTCC 21  
|||||  
Db 788 TCCGATGATGATTTACTCC 808

## RESULT 4

AA199682\_12  
Continuation (13 of 45) of AA199682 from base 1200001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682

WP	Fragment Name	Begin	End
WP	AA199682_00	1	110000
WP	AA199682_01	100001	210000
WP	AA199682_02	200001	310000
WP	AA199682_03	300001	410000
WP	AA199682_04	400001	510000
WP	AA199682_05	500001	610000
WP	AA199682_06	600001	710000
WP	AA199682_07	700001	810000
WP	AA199682_08	800001	910000
WP	AA199682_09	900001	1010000
WP	AA199682_10	1000001	1110000
WP	AA199682_11	1100001	1210000
WP	AA199682_12	1200001	1310000
WP	AA199682_13	1300001	1410000
WP	AA199682_14	1400001	1510000
WP	AA199682_15	1500001	1610000
WP	AA199682_16	1600001	1710000
WP	AA199682_17	1700001	1810000
WP	AA199682_18	1800001	1910000
WP	AA199682_19	1900001	2010000
WP	AA199682_20	2000001	2110000
WP	AA199682_21	2100001	2210000
WP	AA199682_22	2200001	2310000
WP	AA199682_23	2300001	2410000
WP	AA199682_24	2400001	2510000
WP	AA199682_25	2500001	2610000
WP	AA199682_26	2600001	2710000
WP	AA199682_27	2700001	2810000
WP	AA199682_28	2800001	2910000
WP	AA199682_29	2900001	3010000
WP	AA199682_30	3000001	3110000
WP	AA199682_31	3100001	3210000
WP	AA199682_32	3200001	3310000
WP	AA199682_33	3300001	3410000
WP	AA199682_34	3400001	3510000
WP	AA199682_35	3500001	3610000
WP	AA199682_36	3600001	3710000
WP	AA199682_37	3700001	3810000
WP	AA199682_38	3800001	3910000
WP	AA199682_39	3900001	4010000
WP	AA199682_40	4000001	4110000
WP	AA199682_41	4100001	4210000
WP	AA199682_42	4200001	4310000
WP	AA199682_43	4300001	4410000
WP	AA199682_44	4400001	4411529

Query Match 100.0%; Score 21; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCCGATGATGATTTACTCC 21  
|||||  
Db 86397 TCCGATGATGATTTACTCC 86417

## RESULT 5

AA199683\_12  
Continuation (13 of 44) of AA199683 from base 1200001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683

WP	Fragment Name	Begin	End
WP	AA199683_00	1	110000
WP	AA199683_01	100001	210000
WP	AA199683_02	200001	310000
WP	AA199683_03	300001	410000
WP	AA199683_04	400001	510000
WP	AA199683_05	500001	610000
WP	AA199683_06	600001	710000
WP	AA199683_07	700001	810000
WP	AA199683_08	800001	910000
WP	AA199683_09	900001	1010000
WP	AA199683_10	1000001	1110000
WP	AA199683_11	1100001	1210000
WP	AA199683_12	1200001	1310000
WP	AA199683_13	1300001	1410000
WP	AA199683_14	1400001	1510000
WP	AA199683_15	1500001	1610000
WP	AA199683_16	1600001	1710000
WP	AA199683_17	1700001	1810000
WP	AA199683_18	1800001	1910000
WP	AA199683_19	1900001	2010000
WP	AA199683_20	2000001	2110000
WP	AA199683_21	2100001	2210000
WP	AA199683_22	2200001	2310000
WP	AA199683_23	2300001	2410000
WP	AA199683_24	2400001	2510000
WP	AA199683_25	2500001	2610000
WP	AA199683_26	2600001	2710000
WP	AA199683_27	2700001	2810000
WP	AA199683_28	2800001	2910000
WP	AA199683_29	2900001	3010000
WP	AA199683_30	3000001	3110000
WP	AA199683_31	3100001	3210000
WP	AA199683_32	3200001	3310000
WP	AA199683_33	3300001	3410000
WP	AA199683_34	3400001	3510000
WP	AA199683_35	3500001	3610000
WP	AA199683_36	3600001	3710000
WP	AA199683_37	3700001	3810000
WP	AA199683_38	3800001	3910000
WP	AA199683_39	3900001	4010000
WP	AA199683_40	4000001	4110000
WP	AA199683_41	4100001	4210000
WP	AA199683_42	4200001	4310000
WP	AA199683_43	4300001	4403765

Query Match 100.0%; Score 21; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCCGATGATGATTTACTCC 21  
|||||  
Db 85866 TCCGATGATGATTTACTCC 85886

RESULT 6  
ACA47360/c  
ID ACA47360 standard; DNA; 1275 BP.

AC ACA47360;

XX 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #29017.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KM drug design; gene.

XX Staphylococcus haemolyticus.

OS

XX

PN W0200277183-A2.

XX	03-0CT-2002.
XX	
PF	21-MAR-2002; 2002WO-US009107.
XX	
XX	21-MAR-2001; 2001US-00815242.
PR	06-SEP-2001; 2001US-00948993.
PR	25-OCT-2001; 2001US-00342923P.
PR	08-FEB-2002; 2002US-00072851.
PR	06-MAR-2002; 2002US-0362699P.
XX	
PA	(EUIT-) ELITRA PHARM INC.
XX	
PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX	
DR	WPI; 2003-029926/02.
XX	P-PSDB; AEU43490.
PT	
PT	New antisense nucleic acids, useful for identifying proteins or screening
PT	for homologous nucleic acids required for cellular proliferation to
PT	isolate candidate molecules for rational drug discovery programs.
PS	
XX	Claim 14; SEQ ID NO 35230; 1766pp; English.
XX	
CC	The invention relates to an isolated nucleic acid comprising any one of
CC	the 6213 antisense sequences given in the specification where expression
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:
CC	(1) a vector comprising a promoter operably linked to the nucleic acid
CC	encoding a polypeptide whose expression is inhibited by the antisense
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC	polypeptide or its fragment whose expression is inhibited by the
CC	antisense nucleic acid; (4) an antibody capable of specifically binding
CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC	proliferation; or the activity of a gene in an operon required for
CC	proliferation; (7) identifying a compound that influences the activity of
CC	the gene product or that has an activity against a biological pathway; (8)
CC	required for proliferation, or that inhibits cellular proliferation; (9)
CC	identifying a gene required for cellular proliferation or the biological
CC	pathway in which a proliferation-regulated gene or its gene product lies
CC	or a gene on which the test compound that inhibits proliferation of an
CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC	compound's activity; (11) a culture comprising strains in which the gene
CC	product is overexpressed or underexpressed; (12) determining the extent
CC	to which each of the strains is present in a culture or collection of
CC	strains; or (13) identifying the target of a compound that inhibits the
CC	proliferation of an organism. The antisense nucleic acids are useful for
CC	identifying proteins or screening for homologous nucleic acids required
CC	for cellular proliferation to isolate candidate molecules for rational
CC	drug discovery programs, or for screening homologous nucleic acids
CC	required for proliferation in cells other than <i>S. aureus</i> , <i>S. typhimurium</i> ,
CC	<i>K. pneumoniae</i> or <i>P. aeruginosa</i> . The present sequence is one of the target
CC	prokaryotic essential genes. Note: The sequence data for this patent did
CC	not form part of the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
XX	Sequence 1275 BP; 461 A; 179 C; 244 G; 391 T; 0 U; 0 Other;
SO	
Query Match	82.9%; Score 17.4; DB 8; Length 1275;
Best Local Similarity	94.7%; Pred. No. 1,1e+02;
Matches 18; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Oy	1 TCCGGATGATGATTACCT 19
Db	TTCCGATGATGATTACCT 342
RESULT 7	
AAA15451	
ID	AAA15451 standard; DNA; 397 BP.
XX	
AC	AAA15451;

04-SEP-2000 (first entry)

DNA sequence whose expression is regulated by presence of opiates.

Opiate; haematopoietic cell; immune deficiency; blood protein disorder; agammaglobulinemia; dysgammaglobulinemia; ataxia telangiectasia; common variable immunodeficiency; DiGeorge Syndrome; HIV infection; HTLV; Human T-cell leukemia; lymphoma virus infection; leucocyte adhesion deficiency syndrome; lymphopenia; hemoglobinuria; phagocyte bactericidal dysfunction; severe combined immunodeficiency; SCID; Wiskott-Aldrich Disorder; anemia; thrombocytopenia; ss.

Mus musculus.

MO200028026-A2.

18-MAY-2000.

19-OCT-1999; 99MO-US024708.

10-NOV-1998; 98US-0107806P.

(DIGIT-) DIGITAL GENE TECHNOLOGIES INC.

Subcliffe JG, Przewlocki R, Quan J, Thomas EA, WPI; 2000-376541/32.

New polynucleotides whose expression is regulated by opiate administration and withdrawal, useful for treating blood protein disorders, ataxia telangiectasia and common variable immunodeficiency.

Claim 3; Page 83; 85pp; English.

AA15419-53 represent polynucleotides whose expression is regulated by opiate administration and withdrawal. The polynucleotides are useful in methods for preventing, treating or ameliorating a medical condition. Diagnostic methods are used to detect mutations in the polynucleotides. They could be used to increase differentiation and proliferation of haematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types haematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, DiGeorge Syndrome, HIV infection, HTLV (Human T-cell leukemia, lymphoma virus)-BLV infection, leucocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria

Sequence 397 BP, 100 A; 95 C; 81 G; 121 T; 0 U; 0 Other;

Query Match 81.0%; Score 17; DB 3; Length 397;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3 CGGATGATGATTACT 19  
|||||  
298 CGGATGATGATTACT 314

RESULT 8  
ADP77343\_03  
Continuation (4 of 20) of ADP77343 from base 300001 (Lactic acid bacteria Lactobacillus  
WP Sequence Split into 20 fragments LOCUS ADP77343 Accession Adp77343  
WP Fragment Name Begin End  
WP ADP77343\_00 1 110000  
WP ADP77343\_01 100001 210000  
WP ADP77343\_02 200001 310000  
WP ADP77343\_03 300001 410000  
WP ADP77343\_04 400001 510000  
WP ADP77343\_05 500001 610000



```
WP ADF77343_06 600001 710000
WP ADF77343_07 700001 810000
WP ADF77343_08 800001 910000
WP ADF77343_09 900001 1010000
WP ADF77343_10 1000001 1110000
WP ADF77343_11 1100001 1210000
WP ADF77343_12 1200001 1310000
WP ADF77343_13 1300001 1410000
WP ADF77343_14 1400001 1510000
WP ADF77343_15 1500001 1610000
WP ADF77343_16 1600001 1710000
WP ADF77343_17 1700001 1810000
WP ADF77343_18 1800001 1910000
WP ADF77343_19 1900001 1983043

Query Match 81.0%; Score 17; DB 10; Length 110000;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCCGGATGATGATTAC 17
DB 12164 TCCGGATGATGATTAC 12180

RESULT 9
ADQ21179/c
ID ADQ21179 standard; DNA; 131239 BP.
XX
AC ADQ21179;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 3999.
XX
KW soft tissue sarcoma; cytosolic; gene therapy; vaccine; screening; human;
XX ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnick A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 3999; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual,
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cytosolic applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX DNA of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.
```

```
XX
SQ Sequence 131239 BP; 35649 A; 28535 C; 29022 G; 38033 T; 0 U; 0 Other;
XX
Query Match 80.0%; Score 16.8; DB 12; Length 131239;
Best Local Similarity 90.0%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 CCGATGATGATTACTCC 21
DB 131217 CCGATGATGATTACTCC 131198

RESULT 10
ID AC192883 standard; DNA; 25 BP.
XX
AC AC192883;
XX
DT 14-OCT-2003 (first entry)
XX
DE Human microarray DNA oligonucleotide SEQ ID NO 92874.
XX
KW EST; ss; probe; expressed sequence tag; microarray; gene expression;
XX genetic variation; diallelic marker; polymorphism; human;
XX cross-species comparison.
XX
OS Homo sapiens.
XX
PN US2003104410-A1.
XX
PD 05-JUN-2003.
XX
PF 15-MAR-2002; 2002US-00098263.
XX
PR 16-MAR-2001; 2001US-0276759P.
XX
PA (AFPM-) AFPMETRIX INC.
XX
PI Miltmann MP;
XX
DR WPI; 2003-567953/53.
XX
PD New array of nucleic acid probes, useful for in situ hybridization, in
PT Northern, Northern or dot-blot hybridization to identify or detect the
PT sequence or specific mutations of any gene.
XX
PS Claim 1; SEQ ID NO 92874; 9pp; English.
XX
XX The invention discloses a microarray comprising a plurality of nucleic
XX acid probes including one of 2,018,500 fully defined sequences, or its
XX perfect match, perfect mismatch, antisense match or antisense mismatch.
XX Also disclosed is a method of gene expression analysis. The array is used
XX in monitoring gene expression levels by hybridization to a DNA library,
XX in analysis of genetic variation or in hybridization of tag-labelled
XX compounds. The nucleic acid probes are specifically designed for analysis
XX of at least one target sequence. The method of analysis comprises
XX hybridizing at least one or more nucleic acids to at least two or more
XX nucleic acid probes and detecting the hybridization. The nucleic acid
XX probes are attached to a solid support. The analysis comprises monitoring
XX gene expression levels, identifying diallelic markers or polymorphisms,
XX or family members of a gene and a cross-species comparison. Each of the
XX nucleic acids further comprises a tag sequence. The array of nucleic acid
XX probes is useful in in situ hybridization, in Southern, Northern or dot-
XX blot hybridization to identify or detect the sequence or specific
XX mutations of any gene, in mapping the 5' termini of mRNA molecules by
XX primer extensions or in screening cDNA or genomic libraries or subclones
XX for additional subclones containing segments of DNA that have been
XX isolated and previously sequenced. The sequence presented is one of the
XX nucleic acid probes incorporated in the microarray. Note: The sequence
XX data for this patent can also be obtained in electronic format directly
XX from USPTO at seqdata.uspto.gov/sequence.html
XX
SQ Sequence 25 BP; 7 A; 7 C; 8 G; 3 T; 0 U; 0 Other;
```

Query Match 77.1%; Score 16.2; DB 9; Length 25;  
Best Local Similarity 85.7%; Fred. NO. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 TCCGATGATGATTACCTCC 21  
22 TCCGATGTTGAGGTACTCC 2  
Db  
RESULT 11  
AAC41729/c  
ID AAC41729 standard; DNA; 268 BP.  
XX  
AC AAC41729;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 32922.  
XX  
KM Hybridisation assay; Genetic mapping; Gene expression control;  
KM protein identification; signal transduction pathway; metabolic pathway;  
KM promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
OS  
PN BP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123546P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131443P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0135621P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 18-JUN-1999; 99US-0139817P.  
PR 21-JUN-1999; 99US-0139899P.  
PR 22-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140685P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142380P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 20-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 02-AUG-1999; 99US-0147038P.  
PR 03-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.

```
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158023P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0158294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161924P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 77.1%; Score 16.2; DB 3; Length 288;
Best Local Similarity 85.7%; Pred No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy 1 TCCGGATGATTTACTCC 21
||||| ||||| |||||
```

```
DB 77 TCCGGTGTGATTCACCTCC 57
#
RESULT 12
ADA58045
ADA58045 standard, cDNA; 304 BP.
XX
AC ADA58045;
XX
DT 20-NOV-2003 (first entry)
XX
DE Maize triose phosphate isomerase EST #199.
XX
KW Plant; ss; EST; expressed sequence tag; transgenic; sucrose pathway;
KW triose phosphate isomerase; fructose 1, 6-bisphosphate aldolase;
KW fructose 1, 6-bisphosphate; fructose 6-phosphate 2-kinase;
KW phosphoglucosylisomerase; vacuolar hydrogen translocating-pyrophosphatase;
KW pyrophosphate-dependent fructolase-6-phosphate phosphotransferase;
KW invertase; sucrose synthase; hexokinase; fructokinase;
KW nucleoside diphosphate kinase-kinase; NDP;
KW glucose-6-phosphate 1-dehydrogenase; phosphoglucomutase; UDP;
KW uridine diphosphate-glucose pyrophosphorylase; maize.
XX
OS Zea mays.
XX
XX US2003135870-A1.
XX
PD 17-JUL-2003.
XX
PF 26-JAN-1999; 99US-00237183.
XX
PR 24-NOV-1997; 97US-0067000P.
PR 09-DEC-1997; 97US-0069472P.
PR 27-JAN-1998; 98US-0072888P.
PR 10-FEB-1998; 98US-0074201P.
PR 10-FEB-1998; 98US-0074280P.
PR 10-FEB-1998; 98US-0074281P.
PR 10-FEB-1998; 98US-0074282P.
PR 12-FEB-1998; 98US-0074565P.
PR 12-FEB-1998; 98US-0074566P.
PR 12-FEB-1998; 98US-0074567P.
PR 12-FEB-1998; 98US-0074789P.
PR 19-FEB-1998; 98US-0075459P.
PR 19-FEB-1998; 98US-0075460P.
PR 19-FEB-1998; 98US-0075461P.
PR 19-FEB-1998; 98US-0075462P.
PR 19-FEB-1998; 98US-0075463P.
PR 19-FEB-1998; 98US-0075464P.
PR 06-MAR-1998; 98US-0076912P.
PR 09-MAR-1998; 98US-0077229P.
PR 09-MAR-1998; 98US-0077230P.
PR 09-MAR-1998; 98US-0077231P.
PR 18-MAR-1998; 98US-0078368P.
PR 07-APR-1998; 98US-0080844P.
PR 27-APR-1998; 98US-0083067P.
PR 29-APR-1998; 98US-0083366P.
PR 29-APR-1998; 98US-0083387P.
PR 29-APR-1998; 98US-0083388P.
PR 29-APR-1998; 98US-0083389P.
PR 29-APR-1998; 98US-0083390P.
PR 13-MAY-1998; 98US-0085222P.
PR 13-MAY-1998; 98US-0085223P.
PR 13-MAY-1998; 98US-0085224P.
PR 21-MAY-1998; 98US-0086183P.
PR 21-MAY-1998; 98US-0086184P.
PR 21-MAY-1998; 98US-0086185P.
PR 21-MAY-1998; 98US-0086186P.
PR 21-MAY-1998; 98US-0086187P.
PR 01-JUN-1998; 98US-0087422P.
PR 16-JUN-1998; 98US-0089524P.
PR 18-JUN-1998; 98US-0089793P.
PR 18-JUN-1998; 98US-0089810P.
```

PR 18-JUN-1998; 98US-0089814P.  
PR 22-JUN-1998; 98US-0090170P.  
PR 26-JUN-1998; 98US-0090928P.  
PR 29-JUN-1998; 98US-0091035P.  
PR 30-JUN-1998; 98US-0091405P.  
PR 08-JUL-1998; 98US-0092036P.  
PR 09-SEP-1998; 98US-0093667P.  
PR 09-SEP-1998; 98US-0093670P.  
PR 09-SEP-1998; 98US-0093697P.  
PR 16-SEP-1998; 98US-0100672P.  
PR 16-SEP-1998; 98US-0100673P.  
PR 16-SEP-1998; 98US-0100674P.  
PR 21-SEP-1998; 98US-0101131P.  
PR 21-SEP-1998; 98US-0101132P.  
PR 22-SEP-1998; 98US-0101343P.  
PR 22-SEP-1998; 98US-0101344P.  
PR 22-SEP-1998; 98US-0101347P.  
PR 22-SEP-1998; 98US-0101508P.  
PR 25-SEP-1998; 98US-0101707P.  
PR 13-OCT-1998; 98US-0104123P.  
PR 13-OCT-1998; 98US-0104124P.  
PR 13-OCT-1998; 98US-0104126P.  
PR 13-OCT-1998; 98US-0104127P.  
PR 13-OCT-1998; 98US-0104128P.  
PR 18-NOV-1998; 98US-0108996P.  
PR 19-NOV-1998; 98US-0109018P.  
PR 24-NOV-1998; 98US-0019912P.  
PR 08-DEC-1998; 98US-0021029P.  
PR 11-DEC-1998; 98US-0111981P.  
PR 22-DEC-1998; 98US-0113224P.  
PR 12-JAN-1999; 99US-00229413.  
XX  
XX (CHEI/) CHEIKH N.  
XX (FISH/) FISHER D K.  
XX (LIU/) LIU J.  
XX  
XX Cheikh N, Fisher DK, Liu J;  
XX  
XX WPI; 2003-688722/65.  
XX  
XX  
XX New maize or soybean enzymes and nucleic acid molecules associated with  
PT the sucrose pathway, useful for genome mapping, gene identification and  
PT analysis, plant breeding, or preparation of constructs for plant gene  
PT expression.  
XX  
XX  
XX Claim 2; Page: 117pp; English.  
XX  
XX The invention relates to a substantially purified nucleic acid molecule  
CC (appearing as ADA57847 - ADA60660 that encodes a maize or soybean enzyme  
CC or its fragment, associated with the sucrose pathway selected from:  
CC triose phosphate isomerase, fructose 1,6-bisphosphate aldolase, fructose  
CC 1,6-bisphosphate, fructose 6-phosphate 2-kinase, phosphoglucose isomerase,  
CC vacuolar hydrotan, translocating-pyrophosphatase, pyrophosphatase-dependent,  
CC fructose-6-phosphate phosphotransferase, invertase, sucrose synthase,  
CC hexokinase, fructokinase, nucleoside diphosphate kinase (NDP)-kinase,  
CC glucose-6-phosphate 1-dehydrogenase, phosphoglucomutase and uridine  
CC diphosphate (UDP)-glucose pyrophosphorylase. Also included are a  
CC substantially purified maize or soybean enzyme (or its fragment) and a  
CC transformed plant having a nucleic acid molecule. Also disclosed are new  
CC are purified antibodies capable of specifically binding to the maize or  
CC soybean enzyme, determining a level or pattern of a plant sucrose pathway  
CC enzyme in a plant cell or plant tissue, determining a mutation in a plant  
CC whose presence is predictive of a mutation affecting the level or pattern  
CC of a plant sucrose pathway enzyme, producing a plant containing an  
CC overexpressed or reduced level of plant sucrose pathway enzyme, reducing  
CC expression of a plant sucrose pathway enzyme in a plant and determining  
CC an association between a polymorphism and a plant trait. The maize or  
CC soybean enzymes and nucleic acid molecules are useful for genome mapping,  
CC gene identification and analysis, plant breeding, or preparation of  
CC constructs for plant gene expression and transgenic plants. The nucleic  
CC acid molecules are also useful as markers or probes. The present sequence  
CC is a maize EST (expressed sequence tag) from a gene encoding a sucrose

CC pathway enzyme. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from USPTO at [seqdata.uspto.gov/sequence.html?docID=20030135870](http://seqdata.uspto.gov/sequence.html?docID=20030135870).  
XX  
XX  
XX Sequence 304 BP, 57 A, 78 C, 76 G, 93 T, 0 U, 0 Other;  
SQ  
Query Match 77.1%; Score 16.2; DB 9; Length 304;  
Best Local Similarity 85.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 1 TCCGATGATGATTTACTCC 21  
Db 219 TCCGATGCTGATTTCCCTCC 239  
RESULT 13  
ADP60670  
ID ADP60670 standard; cDNA, 304 BP.  
XX  
XX  
XX ADP60670;  
AC  
XX 09-SEP-2004 (first entry)  
DT  
XX  
XX  
XX Maize carbon assimilation pathway enzyme cDNA #1812.  
DE  
XX  
XX Carbon assimilation pathway enzyme; gene; ss; maize; corn;  
XX ribulose-bisphosphate carboxylase; phosphoglycerate kinase;  
XX glyceraldehyde 3-phosphate dehydrogenase; triose phosphate isomerase;  
XX aldolase; fructose-1,6-bisphosphate; transketolase;  
XX sedoheptulose-1,7-bisphosphatase; D-ribulose-5-phosphate-3-epimerase;  
XX ribose-5-phosphate isomerase; ribose-5-phosphate kinase;  
XX phosphoenolpyruvate carboxylase; NADP-dependent malate dehydrogenase;  
XX aspartate aminotransferase; alanine aminotransferase;  
XX NADP-dependent malic enzyme;  
XX PEP carboxykinase; pyruvate; phosphate dikinase; pyrophosphatase;  
XX plant metabolic pathway; plant breeding.  
XX  
XX  
XX Zea mays.  
OS  
XX  
XX US2004116682-A1.  
XX  
XX  
XX 17-JUN-2004.  
XX  
XX  
XX 16-NOV-2001; 2001US-00987899.  
XX  
XX  
XX 06-MAR-1998; 98US-0076912P.  
XX 04-MAR-1999; 99US-00262979.  
XX  
XX  
XX (CHEI/) CHEIKH N.  
XX (MILL/) MILLER P W.  
XX (OCON/) O'CONNELL K M.  
XX (LIU/) LIU J.  
XX  
XX Cheikh N, Miller PW, O'connell KM, Liu J;  
XX  
XX WPI; 2004-498291/47.  
XX  
XX  
XX New substantially purified nucleic acid molecule encoding a maize or  
PT soybean carbon assimilation pathway enzyme, useful for gene mapping, gene  
PT identification and analysis, plant breeding, and preparation of  
PT constructs.  
XX  
XX  
XX Claim 2; SEQ ID NO 3739; 196pp; English.  
XX  
XX The invention relates to a substantially purified nucleic acid molecule  
CC that encodes a maize or soybean carbon assimilation pathway enzyme or its  
CC fragment. The maize or soybean carbon assimilation pathway enzyme or its  
CC fragment is selected from ribulose-bisphosphate carboxylase,  
CC phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, triose  
CC phosphate isomerase, aldolase, fructose-1,6-bisphosphatase,  
CC transketolase, sedoheptulose-1,7-bisphosphatase, D-ribulose-5-phosphate-3-  
CC epimerase, ribose-5-phosphate isomerase, ribose-5-phosphate kinase,  
CC phosphoenolpyruvate carboxylase, NADP-dependent malate dehydrogenase,

CC aspartate aminotransferase, alanine aminotransferase, NADP-dependent  
CC malic enzyme, NAD-dependent malic enzyme, PEP carboxykinase, pyruvate,  
CC phosphate dikinase and pyrophosphatase. The invention also relates to a  
CC substantially purified antibody or its fragment which is capable of  
CC specifically binding to a specific maize or soybean carbon assimilation  
CC pathway enzyme or its fragment, a transformed plant having a nucleic acid  
CC molecule comprising an exogenous promoter region which functions in a  
CC plant cell to cause the production of an mRNA molecule, and a method of  
CC determining a level or pattern in a plant cell of a carbon assimilation  
CC pathway enzyme in a plant metabolic pathway. The methods and compositions  
CC of the invention are useful for gene mapping, gene identification and  
CC analysis, plant breeding and preparation of constructs for use in plant  
CC gene expression and transgenic plants. This sequence represents cDNA  
CC encoding a maize carbon assimilation pathway enzyme of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html.  
CC  
CC  
SQ Sequence 304 BP; 57 A; 78 C; 76 G; 93 T; 0 U; 0 Other;  
  
Query Match 77 1%; Score 16.2; DB 12; Length 304;  
Beet Local Similarity 85.7%; Pred. No. 3,6e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 TCCGGATGATGATTACCTCC 21  
| ||||| ||||| ||||| |||||  
Db 219 TCCGGATGCTGATTTCCTCC 239  
  
RESULT 14  
ID ADA57946  
ADA57946 standard; cDNA; 316 BP.  
XX  
AC ADA57946;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Maize triose phosphate isomerase EST #100.  
XX  
KW Plant; ss; EST; expressed sequence tag; transgenic; sucrose pathway;  
KW triose phosphate isomerase; fructose 1, 6-bisphosphate aldolase;  
KW fructose 1, 6-bisphosphate; fructose 6-phosphate 2-kinase;  
KW phosphoglucosyltransferase; vacuolar hydrogen translocating-pyrophosphatase;  
KW pyrophosphate-dependent fructose-6-phosphate phosphotransferase;  
KW invertase; sucrose synthase; hexokinase; fructokinase;  
KW nucleoside diphosphate kinase-kinase; NDP;  
KW glucose-6-phosphate 1-dehydrogenase; phosphoglucosyltransferase; UDP;  
KW uridine diphosphate-glucose pyrophosphorylase; maize.  
XX  
OS Zea mays.  
XX  
PN US2003135870-A1.  
XX  
PD 17-JUL-2003.  
XX  
PF 26-JAN-1999; 99US-00237183.  
XX  
PR 24-NOV-1997; 97US-0067000P.  
PR 09-DEC-1997; 97US-0069472P.  
PR 27-JAN-1998; 98US-0072888P.  
PR 10-FEB-1998; 98US-0074201P.  
PR 10-FEB-1998; 98US-0074280P.  
PR 10-FEB-1998; 98US-0074281P.  
PR 10-FEB-1998; 98US-0074282P.  
PR 12-FEB-1998; 98US-0074565P.  
PR 12-FEB-1998; 98US-0074566P.  
PR 12-FEB-1998; 98US-0074567P.  
PR 19-FEB-1998; 98US-0074789P.  
PR 19-FEB-1998; 98US-0075459P.  
PR 19-FEB-1998; 98US-0075460P.  
PR 19-FEB-1998; 98US-0075461P.  
PR 19-FEB-1998; 98US-0075462P.  
PR 19-FEB-1998; 98US-0075463P.

PR 19-FEB-1998; 98US-0075464P.  
PR 06-MAR-1998; 98US-0076912P.  
PR 09-MAR-1998; 98US-0077222P.  
PR 09-MAR-1998; 98US-0077230P.  
PR 09-MAR-1998; 98US-0077231P.  
PR 18-MAR-1998; 98US-0078368P.  
PR 07-APR-1998; 98US-0080844P.  
PR 27-APR-1998; 98US-0083067P.  
PR 29-APR-1998; 98US-0083386P.  
PR 29-APR-1998; 98US-0083387P.  
PR 29-APR-1998; 98US-0083388P.  
PR 29-APR-1998; 98US-0083389P.  
PR 29-APR-1998; 98US-0083390P.  
PR 13-MAY-1998; 98US-0085223P.  
PR 13-MAY-1998; 98US-0085223P.  
PR 13-MAY-1998; 98US-0085224P.  
PR 13-MAY-1998; 98US-0086183P.  
PR 21-MAY-1998; 98US-0086184P.  
PR 21-MAY-1998; 98US-0086185P.  
PR 21-MAY-1998; 98US-0086186P.  
PR 21-MAY-1998; 98US-0086187P.  
PR 21-MAY-1998; 98US-0086188P.  
PR 01-JUN-1998; 98US-0087422P.  
PR 16-JUN-1998; 98US-0089524P.  
PR 18-JUN-1998; 98US-0089793P.  
PR 18-JUN-1998; 98US-0089810P.  
PR 18-JUN-1998; 98US-0089811P.  
PR 22-JUN-1998; 98US-0090170P.  
PR 26-JUN-1998; 98US-0090928P.  
PR 29-JUN-1998; 98US-0091035P.  
PR 30-JUN-1998; 98US-0091405P.  
PR 08-JUL-1998; 98US-0092036P.  
PR 09-SEP-1998; 98US-0093667P.  
PR 09-SEP-1998; 98US-0093670P.  
PR 09-SEP-1998; 98US-0093671P.  
PR 16-SEP-1998; 98US-0100672P.  
PR 16-SEP-1998; 98US-0100673P.  
PR 16-SEP-1998; 98US-0100674P.  
PR 21-SEP-1998; 98US-0101130P.  
PR 21-SEP-1998; 98US-0101131P.  
PR 21-SEP-1998; 98US-0101132P.  
PR 22-SEP-1998; 98US-0101343P.  
PR 22-SEP-1998; 98US-0101344P.  
PR 22-SEP-1998; 98US-0101347P.  
PR 22-SEP-1998; 98US-0101506P.  
PR 25-SEP-1998; 98US-0101707P.  
PR 13-OCT-1998; 98US-0104123P.  
PR 13-OCT-1998; 98US-0104124P.  
PR 13-OCT-1998; 98US-0104126P.  
PR 13-OCT-1998; 98US-0104127P.  
PR 13-OCT-1998; 98US-0104128P.  
PR 13-OCT-1998; 98US-0104129P.  
PR 18-NOV-1998; 98US-0108996P.  
PR 19-NOV-1998; 98US-0109018P.  
PR 24-NOV-1998; 98US-0019912P.  
PR 08-DEC-1998; 98US-0021029P.  
PR 11-DEC-1998; 98US-0111981P.  
PR 22-DEC-1998; 98US-0113224P.  
PR 12-JAN-1999; 99US-00229413.  
XX  
XX (CHEIKH) CHEIKH N.  
PA (FISHER) FISHER D K.  
PA (LIU/J) LIU J.  
PI Cheikh N, Fisher DK, Liu J;  
XX  
XX WPI; 2003-688722/65.  
PT New maize or soybean enzymes and nucleic acid molecules associated with  
PT the sucrose pathway, useful for genome mapping, gene identification and  
PT analysis, plant breeding, or preparation of constructs for plant gene  
PT expression.  
XX  
XX Claim 2; Page; 117pp; English.

XX The invention relates to a substantially purified nucleic acid molecule  
CC (appearing as AD657847 - ADA60660 that encodes a maize or soybean enzyme  
CC or its fragment, associated with the sucrose pathway selected from:  
CC triose phosphate isomerase, fructose 1,6-bisphosphate aldolase, fructose  
CC 1,6-bisphosphate isomerase, fructose 6-phosphate 2-kinase, phosphoglucose  
CC isomerase, fructose 6-phosphate phosphatase, pyrophosphate-dependent  
CC fructose-6-phosphate phosphotransferase, invertase, sucrose synthase,  
CC hexokinase, fructokinase, nucleoside diphosphate kinase (NDP)-kinase,  
CC glucose-6-phosphate 1-dehydrogenase, phosphoglucomutase and uridine  
CC diphosphate (UDP)-glucose pyrophosphorylase. Also included are a  
CC substantially purified maize or soybean enzyme (or its fragment) and a  
CC transformed plant having a nucleic acid molecule. Also disclosed as new  
CC are purified antibodies capable of specifically binding to the maize or  
CC soybean enzyme, determining a level or pattern of a plant sucrose pathway  
CC enzyme in a plant cell or plant tissue, determining a mutation in a plant  
CC whose presence is predictive of a mutation affecting the level or pattern  
CC of a plant sucrose pathway enzyme, producing a plant containing an  
CC overexpressed or reduced level of plant sucrose pathway enzyme, reducing  
CC expression of a plant sucrose pathway enzyme in a plant and determining  
CC an association between a polymorphism and a plant trait. The maize or  
CC soybean enzymes and nucleic acid molecules are useful for genome mapping,  
CC gene identification and analysis, plant breeding, or preparation of  
CC constructs for plant gene expression and transgenic plants. The nucleic  
CC acid molecules are also useful as markers or probes. The present sequence  
CC is a maize EST (expressed sequence tag) from a gene encoding a sucrose  
CC pathway enzyme. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from USPTO at [seqdata.uspto.gov/sequence.html?docid=20030135870](http://seqdata.uspto.gov/sequence.html?docid=20030135870).  
XX  
SQ Sequence 316 BP; 66 A; 81 C; 80 G; 89 T; 0 U; 0 Other;  
Query Match 77.1%; Score 16.2; DB 9; Length 316;  
Best Local Similarity 85.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 TCCGGATGATGATTACTCTCC 21  
Db 288 TGGCGATGTCGATTTCCTCC 308  
RESULT 15  
ADP60571  
ID ADP60571 standard; cDNA; 316 BP.  
XX  
AC ADP60571;  
XX  
XX 09-SEP-2004 (first entry)  
DT  
XX  
XX  
DE Maize carbon assimilation pathway enzyme cDNA #1713.  
XX  
XX Carbon assimilation pathway enzyme; gene; ss; maize; corn;  
KW ribulose-bisphosphate carboxylase; phosphoglycerate kinase;  
KW glyceraldehyde 3-phosphate dehydrogenase; triose phosphate isomerase;  
KW aldolase; fructose-1,6-bisphosphate; transketolase;  
KW sedoheptulose-1,7-bisphosphatase; D-ribulose-5-phosphate-3-epimerase;  
KW ribose-5-phosphate isomerase; ribose-5-phosphate kinase;  
KW phosphoenolpyruvate carboxylase; NADP-dependent malate dehydrogenase;  
KW aspartate aminotransferase; alanine aminotransferase;  
KW NADP-dependent malic enzyme; NAD-dependent malic enzyme;  
KW PEP carboxykinase; pyruvate; phosphate dikinase; pyrophosphatase;  
KW plant metabolic pathway; plant breeding.  
XX  
OS Zea mays.  
XX  
XX US2004116682-A1.  
XX  
XX 17-JUN-2004.  
XX  
XX 16-NOV-2001; 2001US-00987899.  
XX  
XX 06-MAR-1998; 98US-0076912P.  
XX  
XX 04-MAR-1999; 99US-00262979.  
XX  
PR

XX  
PA (CHEI/) CHEIKH N.  
PA (MILL/) MILLER P W.  
PA (OCON/) O'CONNELL K M.  
PA (LIU/) LIU J.  
XX Cheikh N, Miller PW, O'Connell KM, Liu J,  
DR WPI, 2004-498291/47.  
XX  
PT New substantially purified nucleic acid molecule encoding a maize or  
PT soybean carbon assimilation pathway enzyme, useful for gene mapping, gene  
PT identification and analysis, plant breeding, and preparation of  
PT constructs.  
XX  
PS Claim 2; SEQ ID NO 3640; 196pp; English.  
XX  
XX The invention relates to a substantially purified nucleic acid molecule  
CC that encodes a maize or soybean carbon assimilation pathway enzyme or its  
CC fragment. The maize or soybean carbon assimilation pathway enzyme or its  
CC fragment is selected from ribulose-bisphosphate carboxylase,  
CC phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, triose  
CC phosphate isomerase, aldolase, fructose-1,6-bisphosphatase, fructose-3  
CC transketolase, sedoheptulose-1,7-bisphosphatase, D-ribulose-5-phosphate-3  
CC -epimerase, ribose-5-phosphate isomerase, ribose-5-phosphate kinase,  
CC phosphoenolpyruvate carboxylase, NADP-dependent malate dehydrogenase,  
CC aspartate aminotransferase, alanine aminotransferase, NADP-dependent  
CC malic enzyme, NAD-dependent malic enzyme, PEP carboxykinase, pyruvate,  
CC phosphate dikinase and pyrophosphatase. The invention also relates to a  
CC substantially purified antibody or its fragment which is capable of  
CC specifically binding to a specific maize or soybean carbon assimilation  
CC pathway enzyme or its fragment, a transformed plant having a nucleic acid  
CC molecule comprising an exogenous promoter region which functions in a  
CC plant cell to cause the production of an mRNA molecule, and a method of  
CC determining a level or pattern in a plant cell of a carbon assimilation  
CC pathway enzyme in a plant metabolic pathway. The methods and compositions  
CC of the invention are useful for gene mapping, gene identification and  
CC analysis, plant breeding and preparation of constructs for use in plant  
CC gene expression and transgenic plants. This sequence represents cDNA  
CC encoding a maize carbon assimilation pathway enzyme of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
XX  
SQ Sequence 316 BP; 66 A; 81 C; 80 G; 89 T; 0 U; 0 Other;  
Query Match 77.1%; Score 16.2; DB 12; Length 316;  
Best Local Similarity 85.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 TCCGGATGATGATTACTCTCC 21  
Db 288 TGGCGATGTCGATTTCCTCC 308  
Search completed: November 7, 2005, 05:40:00  
Job time : 254.17 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 19:19:37 ; Search time 71.266 Seconds  
(without alignments)  
482.163 Million cell updates/sec

Title: US-10-777-131A-1

Perfect score: 21

Sequence: 1 tccgcatgatcttactctc 21

Scoring table: IDENTITY\_NUC

Searched: Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCtUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
2	21	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
3	16.8	80.0	4308	4 US-09-614-221A-599	Sequence 599, App
4	16.2	77.1	1131	4 US-09-328-352-1902	Sequence 14902, A
5	16.2	77.1	44019	4 US-09-949-016-16902	Sequence 95, Appl
6	16	76.2	3510	4 US-09-265-585C-95	Sequence 12128, A
7	15.8	75.2	666	4 US-09-248-796A-12128	Sequence 181, App
8	15.4	73.3	415	4 US-09-220-132-181	Sequence 18981, A
9	15.4	73.3	601	4 US-09-949-016-18982	Sequence 18983, A
10	15.4	73.3	601	4 US-09-949-016-18982	Sequence 18983, A
11	15.4	73.3	601	4 US-09-949-016-18983	Sequence 47061, A
12	15.4	73.3	601	4 US-09-949-016-47061	Sequence 47062, A
13	15.4	73.3	601	4 US-09-949-016-47062	Sequence 2549, Ap
14	15.4	73.3	750	4 US-09-543-681A-2549	Sequence 131, App
15	15.4	73.3	1173	3 US-09-561-332-1	Sequence 777, App
16	15.4	73.3	1337	4 US-09-220-132-131	Sequence 13118, A
17	15.4	73.3	2742	4 US-09-799-451-777	Sequence 13119, A
18	15.4	73.3	107800	4 US-09-949-016-13118	Sequence 11809, A
19	15.4	73.3	116425	4 US-09-949-016-11809	Sequence 12201, A
20	15.4	73.3	227390	4 US-09-949-016-12201	Sequence 11365, A
21	15.4	73.3	227391	4 US-09-949-016-13365	Sequence 11368, A
22	15.4	73.3	636591	4 US-09-949-016-13388	Sequence 1211, Ap
23	15.4	73.3	636591	4 US-09-513-999C-1211	Sequence 13712, A
24	15.2	72.4	346	4 US-09-621-976-13712	Sequence 2121, Ap
25	15.2	72.4	358	4 US-09-621-976-13712	
26	15.2	72.4	407	4 US-09-621-976-2121	
27	15.2	72.4	407	4 US-09-621-976-2121	

28	15.2	72.4	413	3 US-09-227-357-71	Sequence 71, Appl
29	15.2	72.4	427	4 US-09-513-999C-10820	Sequence 10820, A
30	15.2	72.4	601	4 US-09-949-016-60865	Sequence 60865, A
31	15.2	72.4	601	4 US-09-949-016-127401	Sequence 127401, A
32	15.2	72.4	2264	2 US-08-149-097D-26	Sequence 26, Appl
33	15.2	72.4	2404	4 US-09-269-386-26	Sequence 26, Appl
34	15.2	72.4	2634	3 US-08-984-386-26	Sequence 26, Appl
35	15.2	72.4	2634	3 US-08-450-562-26	Sequence 26, Appl
36	15.2	72.4	2634	3 US-08-984-709A-26	Sequence 26, Appl
37	15.2	72.4	2634	3 US-08-450-272-26	Sequence 26, Appl
38	15.2	72.4	2634	3 US-08-450-273-26	Sequence 26, Appl
39	15.2	72.4	2712	3 US-08-949-386-38	Sequence 38, Appl
40	15.2	72.4	2712	3 US-08-450-562-38	Sequence 38, Appl
41	15.2	72.4	2712	3 US-08-984-709A-38	Sequence 38, Appl
42	15.2	72.4	2712	3 US-08-450-272-38	Sequence 38, Appl
43	15.2	72.4	2909	3 US-08-450-273-38	Sequence 38, Appl
44	15.2	72.4	2909	3 US-08-104-158-1	Sequence 1, Appl
45	15.2	72.4	2909	4 US-09-609-040-1	Sequence 1, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-103-840A-2
Sequence 2, 6294328
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
Query Match 100.0%; Score 21; DB 3; Length 4403765;
Best local similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 TCCGATGATGATTTACTCTC 21
DB 1285866 TCCGATGATGATTTACTCTC 1285866
RESULT 2
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
```



```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match          100.0%; Score 21; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TCCGATGATGATTACTCTCC 21
Db      1286397 TCCGATGATGATTACTCTCC 1286417
```

```
RESULT 3
US-09-614-221A-599
; Sequence 599, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunamanda, Balasubojini
; APPLICANT: Yu, Jaehyuk
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516_075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 599
; LENGTH: 4308
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-599
```

```
Query Match          80.0%; Score 16.8; DB 4; Length 4308;
Best Local Similarity 90.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 CCGATGATGATTACTCTCC 21
Db      1018 CCGAAGATGATTACCCACC 1037
```

```
RESULT 4
US-09-328-352-1902
; Sequence 1902, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1902
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1902
```

```
Query Match          77.1%; Score 16.2; DB 4; Length 1131;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 TCCGATGATGATTACTCTCC 21
Db      1098 TTCAGATGATGATTACCGCC 1118
```

```
RESULT 5
US-09-949-016-14902/C
; Sequence 14902, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14902
; LENGTH: 44019
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc.feature
; LOCATION: (1) - (44019)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14902
```

```
Query Match          77.1%; Score 16.2; DB 4; Length 44019;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 TCCGATGATGATTACTCTCC 21
Db      14173 TCCAGATGATGATTACTCTCC 14153
```

```
RESULT 6
US-09-265-585C-95/C
; Sequence 95, Application US/09265585C
; Patent No. 6809234
; GENERAL INFORMATION:
; APPLICANT: Benfey, Phillip N.
; APPLICANT: Di Laurentio, Laura
; APPLICANT: Wypocka-Diller, Joanna
; APPLICANT: Malamy, Jocelyn E.
; APPLICANT: Pysh, Leonard
; APPLICANT: Helariutta, Yrjo
; APPLICANT: Bruce, Weesley
; APPLICANT: Lim, Jun
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
; FILE REFERENCE: 5914-066
; CURRENT APPLICATION NUMBER: US/09/265,585C
; CURRENT FILING DATE: 1999-03-10
; PRIOR APPLICATION NUMBER: 08/842,445
; PRIOR FILING DATE: 1997-04-24
; PRIOR APPLICATION NUMBER: 08/638,617
; PRIOR FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 152
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 95
; LENGTH: 3510
; TYPE: DNA
; ORGANISM: Zea mays
```

```
FEATURE:
; NAME/KEY: CDS
; LOCATION: (293)..(1855)
; NAME/KEY: CDS
; LOCATION: (2703)..(3143)
US-09-265-585C-95
```

Query Match 76.2%; Score 16; DB 4; Length 3510;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGGATGATGATTCTTA 16  
|||  
Db 2386 TCCGGATGATGATTCTTA 2371

RESULT 7  
US-09-248-796A-12128  
; Sequence 12128, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 12128  
; LENGTH: 666  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-09-248-796A-12128

Query Match 75.2%; Score 15.8; DB 4; Length 666;  
Best Local Similarity 89.5%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCCGGATGATGATTCTTACT 19  
|||  
Db 219 TCCGTGATGCTGATTCTTACT 237

RESULT 8  
US-09-220-132-181/c  
; Sequence 181, Application US/09220132  
; Patent No. 6506607  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
; FILE REFERENCE: 07334-074001  
; CURRENT APPLICATION NUMBER: US/09/220,132  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: US 60/079,303  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: US 60/068,821  
; PRIOR FILING DATE: 1997-12-24  
; NUMBER OF SEQ ID NOS: 191  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 181  
; LENGTH: 415  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(415)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-220-132-181

Query Match 73.3%; Score 15.4; DB 4; Length 415;  
Best Local Similarity 94.1%; Pred. No. 2.3e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GATGATGATTTACTCTCC 21

Db 283 GATGATGATTTACTCTCC 267

RESULT 9  
US-09-949-016-18981/c  
; Sequence 18981, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18981  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-18981

Query Match 73.3%; Score 15.4; DB 4; Length 601;  
Best Local Similarity 94.1%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CGGATGATGATTACTT 19  
|||  
Db 389 CGGCTGATGATTACTT 373

RESULT 10  
US-09-949-016-18982/c  
; Sequence 18982, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18982  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-18982

Query Match 73.3%; Score 15.4; DB 4; Length 601;  
Best Local Similarity 94.1%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CGGATGATGATTACTT 19  
|||  
Db 394 CGGCTGATGATTACTT 378

RESULT 11

US-09-949-016-18983/c  
; Sequence 18983, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18983  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-18983

Query Match 73.3%; Score 15.4; DB 4; Length 601;  
Best Local Similarity 94.1%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CGGATGATGATTACT 19  
|||  
Db 486 CGGCTGATGATTACT 470

RESULT 12  
US-09-949-016-47061/c  
; Sequence 47061, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 47061  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-47061

Query Match 73.3%; Score 15.4; DB 4; Length 601;  
Best Local Similarity 94.1%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CGGATGATGATTACT 19  
|||  
Db 389 CGGCTGATGATTACT 373

RESULT 13  
US-09-949-016-47062/c  
; Sequence 47062, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 47062  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-47062

Query Match 73.3%; Score 15.4; DB 4; Length 601;  
Best Local Similarity 94.1%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CGGATGATGATTACT 19  
|||  
Db 394 CGGCTGATGATTACT 378

RESULT 14  
US-09-949-016-47063/c  
; Sequence 47063, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 47063  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-47063

Query Match 73.3%; Score 15.4; DB 4; Length 601;  
Best Local Similarity 94.1%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CGGATGATGATTACT 19  
|||  
Db 486 CGGCTGATGATTACT 470

RESULT 15  
US-09-543-681A-2549  
; Sequence 2549, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706  
 ; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 8344  
 ; SEQ ID NO 2549  
 ; LENGTH: 750  
 ; TYPE: DNA  
 ; ORGANISM: Proteus mirabilis  
 US-09-543-681A-2549

Query Match 73.3%; Score 15.4; DB 4; Length 750;  
 Best Local Similarity 94.1%; Pred. No. 2.6e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GGATGATGATTACCTC 20  
 |||||  
 Db 57 GGATGATGATTACCTC 73

Search completed: November 6, 2005, 23:09:34  
 Job time : 92.266 secs

**This Page Blank (user)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 20:56:42 ; Search time 523.66 Seconds  
(without alignments)  
331.640 Million cell updates/sec

Title: US-10-777-131a-1

Perfect score: 21  
Sequence: 1 tccgcgatgcattactctcc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 20: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 21: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 22: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 23: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 24: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 25: /cgn2\_6/ptodata/2/pubpna/US11\_PUBCOMB.seq.\*
- 26: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 27: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 28: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	US-10-216-817-1	Sequence 1, Appli
2	21	100.0	21	US-10-777-131a-1	Sequence 1, Appli
3	21	100.0	306	US-10-755-415-92	Sequence 92, Appli
4	21	100.0	823	US-10-216-817-19	Sequence 19, Appli
5	21	100.0	823	US-10-777-131a-19	Sequence 19, Appli

C	6	17.8	84.8	669	21	US-10-425-115-10736	Sequence 10736, A
	7	17.8	84.8	4509	20	US-10-437-963-22127	Sequence 22127, A
	8	17.8	84.8	4596	20	US-10-437-963-22129	Sequence 22129, A
C	9	17.4	82.9	4303	19	US-10-424-599-89548	Sequence 89548, A
C	10	17.4	82.9	1275	18	US-10-424-599-87436	Sequence 87436, A
C	11	17.4	82.9	1275	18	US-10-282-122A-35230	Sequence 35230, A
C	12	16.8	80.0	397	20	US-10-437-963-63692	Sequence 63692, A
	13	16.8	80.0	1821	20	US-10-437-963-52282	Sequence 52282, A
	14	16.8	80.0	2181	20	US-10-437-963-52283	Sequence 52283, A
	15	16.8	80.0	4308	20	US-10-793-639-599	Sequence 599, A
C	16	16.8	80.0	131239	21	US-10-723-860-3999	Sequence 3999, A
C	17	16.8	80.0	131239	24	US-10-756-149-3963	Sequence 3963, A
	18	16.4	78.1	581	14	US-10-027-632-286839	Sequence 286839, A
	19	16.4	78.1	581	18	US-10-027-632-286839	Sequence 286839, A
C	20	16.4	78.1	2250	20	US-10-437-963-100004	Sequence 100004, A
	21	16.2	77.1	25	16	US-10-098-2638-92874	Sequence 92874, A
	22	16.2	77.1	249	21	US-10-425-115-84686	Sequence 84686, A
	23	16.2	77.1	304	10	US-09-237-183A-199	Sequence 199, A
	24	16.2	77.1	304	11	US-09-237-183A-197	Sequence 197, A
	25	16.2	77.1	316	10	US-09-237-183A-100	Sequence 100, A
	26	16.2	77.1	316	11	US-09-237-183A-101	Sequence 101, A
	27	16.2	77.1	325	10	US-09-237-183A-101	Sequence 101, A
	28	16.2	77.1	325	11	US-09-237-183A-101	Sequence 101, A
	29	16.2	77.1	393	19	US-10-425-114-2846	Sequence 2846, A
	30	16.2	77.1	398	10	US-09-237-183A-197	Sequence 197, A
	31	16.2	77.1	398	11	US-09-237-183A-197	Sequence 197, A
	32	16.2	77.1	483	13	US-09-925-065A-219459	Sequence 219459, A
	33	16.2	77.1	483	13	US-09-925-065A-219460	Sequence 219460, A
C	34	16.2	77.1	582	21	US-10-425-115-30270	Sequence 30270, A
C	35	16.2	77.1	592	24	US-10-972-079-87898	Sequence 87898, A
	36	16.2	77.1	600	24	US-10-972-079-87897	Sequence 87897, A
	37	16.2	77.1	645	14	US-10-027-632-280976	Sequence 280976, A
	38	16.2	77.1	645	18	US-10-027-632-280976	Sequence 280976, A
	39	16.2	77.1	656	14	US-10-027-632-212304	Sequence 212304, A
	40	16.2	77.1	656	14	US-10-027-632-212305	Sequence 212305, A
	41	16.2	77.1	656	14	US-10-027-632-212306	Sequence 212306, A
	42	16.2	77.1	656	18	US-10-027-632-212306	Sequence 212306, A
	43	16.2	77.1	656	18	US-10-027-632-212305	Sequence 212305, A
	44	16.2	77.1	656	18	US-10-027-632-212306	Sequence 212306, A
	45	16.2	77.1	726	19	US-10-425-114-6761	Sequence 6761, A

#### ALIGNMENTS

RESULT 1  
US-10-216-817-1  
; Sequence 1, Application US/10216817  
; Publication No. US20030129619A1  
GENERAL INFORMATION:  
APPLICANT: GICQUEL, BRIGITTE  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
FILE REFERENCE: 03495.0233-00000  
CURRENT APPLICATION NUMBER: US/10/216,817  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 60/311,824  
PRIOR FILING DATE: 2001-08-14  
PRIOR APPLICATION NUMBER: 60/313,523  
PRIOR FILING DATE: 2001-08-21  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 21  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURES:  
OTHER INFORMATION: Description of Artificial Sequence: Primer  
US-10-216-817-1  
Query Match 100.0%; Score 21; DB 16; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21  
Db 1 TCCGATGATGATTACTCC 21

RESULT 2  
US-10-777-131A-1

; Sequence 1, Application US/10777131A  
; Publication No. US2005026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; TITLE OF INVENTION: IN GENES OF THE MULTIFAMILY  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT APPLICATION NUMBER: US/10/777,131A  
; PRIOR FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 1  
; LENGTH: 21  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer  
US-10-777-131A-1

Query Match 100.0%; Score 21; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21  
Db 1 TCCGATGATGATTACTCC 21

RESULT 3

; Sequence 92, Application US/10755415  
; Publication No. US20050136480A1  
; GENERAL INFORMATION:  
; APPLICANT: BRAHMACHARI, SAMIR KUMAR  
; APPLICANT: DASH, DEBASIS  
; APPLICANT: SHARMA, RAMAKANT  
; APPLICANT: MAHESHWARI, JITENDRA KUMAR  
; TITLE OF INVENTION: A COMPUTER BASED VERSATILE METHOD FOR IDENTIFYING PROTEIN CODING  
; TITLE OF INVENTION: DNA SEQUENCES USEFUL AS DRUG TARGETS  
; FILE REFERENCE: 026033-00029  
; CURRENT APPLICATION NUMBER: US/10/755,415  
; PRIOR FILING DATE: 2004-01-13  
; PRIOR APPLICATION NUMBER: 10/727,989  
; PRIOR FILING DATE: 2003-12-05  
; NUMBER OF SEQ ID NOS: 373  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 92  
; LENGTH: 306  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-755-415-92

Query Match 100.0%; Score 21; DB 24; Length 306;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21

Db 191 TCCGATGATGATTACTCC 171

RESULT 4  
US-10-216-817-19

; Sequence 19, Application US/10216817  
; Publication No. US20030129619A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; TITLE OF INVENTION: IN GENES OF THE MULTIFAMILY  
; FILE REFERENCE: 03495.0233-00000  
; CURRENT APPLICATION NUMBER: US/10/216,817  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 823  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-216-817-19

Query Match 100.0%; Score 21; DB 16; Length 823;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21  
Db 5 TCCGATGATGATTACTCC 25

RESULT 5  
US-10-777-131A-19

; Sequence 19, Application US/10777131A  
; Publication No. US20050026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; TITLE OF INVENTION: IN GENES OF THE MULTIFAMILY  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT APPLICATION NUMBER: US/10/777,131A  
; PRIOR FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 19  
; LENGTH: 823  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-777-131A-19

Query Match 100.0%; Score 21; DB 22; Length 823;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21  
Db 5 TCCGATGATGATTACTCC 25

RESULT 6



```
US-10-425-115-10736/c
; Sequence 10736, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 10736
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_109792C.1
US-10-425-115-10736

Query Match      84.8% Score 17.8; DB 21; Length 669;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 TCCGATGATGATTTACCTCC 21
      |||||
Db      566 TCCGATGATGATTTACCTCC 546

RESULT 7
US-10-437-963-22127
; Sequence 22127, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 22127
; LENGTH: 4509
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_27332C.1
US-10-437-963-22127

Query Match      84.8% Score 17.8; DB 20; Length 4509;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 TCCGATGATGATTTACCTCC 21
      |||||
Db      165 TCCGATGATGATTTACCAACC 185

RESULT 8
US-10-437-963-22129
; Sequence 22129, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 89548
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51872C.1
US-10-424-599-89548

Query Match      82.9% Score 17.4; DB 19; Length 340;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 CCGATGATGATTTACCTC 20
      |||||
Db      181 CCGATGATGATTTACCC 163

RESULT 9
US-10-424-599-89548/c
; Sequence 89548, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 89548
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51872C.1
US-10-424-599-89548

Query Match      82.9% Score 17.4; DB 19; Length 340;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 CCGATGATGATTTACCTC 20
      |||||
Db      181 CCGATGATGATTTACCC 163

RESULT 10
US-10-424-599-87436/c
; Sequence 87436, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 89548
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51872C.1
US-10-424-599-89548
```

```
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 87436
LENGTH: 403
TYPE: DNA
ORGANISM: glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_49965C.1
US-10-424-599-87436

Query Match      82.9%; Score 17.4; DB 19; Length 403;
Best Local Similarity 94.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 CCGATGATGATTACCTC 20
Db      235 CCGATGATGATTACCCC 217

RESULT 11
US-10-282-122A-35230/c
Sequence 35230, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyckind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35230
LENGTH: 1275
TYPE: DNA
ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-35230

Query Match      82.9%; Score 17.4; DB 18; Length 1275;
Best Local Similarity 94.7%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 TCCGATGATGATTACCT 19
Db      360 TTCGATGATGATTACCT 342

RESULT 12
US-10-437-963-63692/c
Sequence 63692, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 63692
LENGTH: 397
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_6490C.1
US-10-437-963-63692

Query Match      80.0%; Score 16.8; DB 20; Length 397;
Best Local Similarity 90.0%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      2 CCGATGATGATTACCTC 21
Db      40 CCGATGATGATTACCCC 21

RESULT 13
US-10-437-963-52292
Sequence 52292, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 52292
LENGTH: 1821
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_54601C.1
US-10-437-963-52292

Query Match      80.0%; Score 16.8; DB 20; Length 1821;
Best Local Similarity 90.0%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

OY 2 CCGATGATGATTACTCTCC 21  
|||  
Db 1207 CCGAAGATGATTACTCTCC 1226

Search completed: November 7, 2005, 05:20:32  
Job time : 525.66 secs

RESULT 14  
US-10-437-963-52293  
; Sequence 52293, Application US/10437963  
; Publication No. US20040123343A1  
GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Bardazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 52293  
; LENGTH: 2181  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_54602C.1  
US-10-437-963-52293

Query Match 80.0%; Score 16.8; DB 20; Length 2181;  
Best Local Similarity 90.0%; Pred. No. 5.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 CCGATGATGATTACTCTCC 21  
|||  
Db 1567 CCGAAGATGATTACTCTCC 1586

RESULT 15  
US-10-793-639-599  
; Sequence 599, Application US/10793639  
; Publication No. US20040199940A1  
GENERAL INFORMATION:  
; APPLICANT: Karunanandaa, Balasubramanian  
; APPLICANT: Yu, Jaehyuk  
; APPLICANT: Kishore, Ganesh M.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED  
; TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM  
; FILE REFERENCE: 16516.075  
; CURRENT APPLICATION NUMBER: US/10/793,639  
; CURRENT FILING DATE: 2004-03-05  
; PRIOR APPLICATION NUMBER: US/09/614,221A  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: US 60/142,981  
; PRIOR FILING DATE: 1999-07-12  
; NUMBER OF SEQ ID NOS: 626  
; SEQ ID NO 599  
; LENGTH: 4308  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
US-10-793-639-599

Query Match 80.0%; Score 16.8; DB 20; Length 4308;  
Best Local Similarity 90.0%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 CCGATGATGATTACTCTCC 21  
|||  
Db 1018 CCGAAGATGATTACTCTCC 1037

*This Page Blank (usp10)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 2100.89 Seconds  
(without alignments)  
380.481 Million cell updates/sec

Title: US-10-777-131A-1

Perfect score: 21

Sequence: 1 tccgagatgatctactcc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 3439544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_hic:  
4: gb\_est3:  
5: gb\_est4:  
6: gb\_est5:  
7: gb\_est6:  
8: gb\_gsa1:  
9: gb\_gsa2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	90.5	773	8	BZ499562 BONMG24TR
2	18.4	87.6	668	5	BP136446
3	18	85.7	764	7	CP439390 EST675735
4	17.8	84.8	948	9	CL965683 OSIFCC012
5	17.8	84.8	983	4	BM472190
6	17.4	82.9	200	1	AV340408
7	17.4	82.9	805	7	CO809020
8	17	81.0	196	6	CA328214
9	17	81.0	289	1	AA254497
10	17	81.0	286	1	AV164857
C 11	17	81.0	320	4	BG060357
C 12	17	81.0	326	2	BEG34382
C 13	17	81.0	337	5	BX525225
C 14	17	81.0	399	2	AM046909
C 15	17	81.0	401	6	BY609421
C 16	17	81.0	414	2	BE991046
C 17	17	81.0	414	2	BE991225
C 18	17	81.0	416	5	BX636140
C 19	17	81.0	417	1	AA085413
C 20	17	81.0	418	7	CF744036
C 21	17	81.0	425	5	BU611514
C 22	17	81.0	466	5	BF582889
C 23	17	81.0	472	4	BG868769
C 24	17	81.0	484	1	AA880291

C 25	17	81.0	524	4	BM241795
C 26	17	81.0	538	7	BR853277
C 27	17	81.0	577	2	CK616236
C 28	17	81.0	623	2	AM550920
C 29	17	81.0	649	1	AF157269
C 30	17	81.0	722	1	AF122567
C 31	17	81.0	857	8	CC435783
C 32	16.8	80.0	193	2	BR461953
C 33	16.8	80.0	256	9	CL221696
C 34	16.8	80.0	263	9	CR097239
C 35	16.8	80.0	394	2	BR432151
C 36	16.8	80.0	462	2	BR435941
C 37	16.8	80.0	486	2	BE460443
C 38	16.8	80.0	486	4	BM405690
C 39	16.8	80.0	500	2	BF051758
C 40	16.8	80.0	507	2	BE435744
C 41	16.8	80.0	523	9	CL326924
C 42	16.8	80.0	527	4	BM535572
C 43	16.8	80.0	538	2	BE460424
C 44	16.8	80.0	541	5	BX332222
C 45	16.8	80.0	593	5	BP505841

#### ALIGNMENTS

RESULT 1  
LOCUS BZ499562 773 bp DNA linear GSS 16-DEC-2002  
DEFINITION BONMG24TR BO.1.6.2 KB tot Brassica oleracea genomic clone BONMG24,  
genomic survey sequence.  
ACCESSION BZ499562  
VERSION BZ499562.1 GI:27014783  
KEYWORDS  
SOURCE  
ORGANISM  
Brassica oleracea  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE  
Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Frazer,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
unpublished (2001)  
COMMENT  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

#### FEATURES

source  
1..773  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO100DH3"  
/db\_xref="taxon:3712"  
/clone="BONMG24"  
/clone\_1lb="BO.1.6.2 KB tot"  
/note="Vector: pHD1. Site 1: BstXI; 1.6-2 kb sheared  
total DNA inserted into pHD1 using BstXI linkers"

#### ORIGIN

Query Match 90.5%; Score 19; DB 8; Length 773;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 2 CCGATGATGATTTACTTC 20  
DB 248 CCGATGATGATTTACTTC 230

```

RESULT 2
LOCUS      Bp136446                      668 bp    mRNA    linear    EST 17-JUL-2003
DEFINITION Bp136446 MAT001 Nicotiana tabacum cDNA clone BY9044, mRNA sequence.
ACCESSION  Bp136446
VERSION     Bp136446.1 GI:32890036
KEYWORDS   EST.
SOURCE      Nicotiana tabacum (common tobacco)
ORGANISM   Nicotiana tabacum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamids; Solanales; Solanaceae; Nicotiana.
REFERENCE  1 (bases 1 to 668)
            Matsunaka, K., Tashiro, G., Horiuchi, T., Demura, T. and Fukuda, H.
            Profiling growth-phase dependent gene expression of tobacco BY-2
            cells by comprehensive microarray analysis
            Unpublished (2003)
JOURNAL    Contact: Ken Matsunaka
            Morphogenesis Research Group
            RIKEN Plant Science Center
            1-7-2 Suenohirocho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9575
            Fax: 81-45-503-9573
            Email: by2@pac.riken.go.jp, URL: http://mrg.psc.riken.go.jp/stc/
            The cDNA library was constructed from mRNA isolated from lag (9 h),
            lag (72 h) and stationary (7 days) old BY-2 cells.
FEATURES   Location/Qualifiers
            source          1..668
                        /organism="Nicotiana tabacum"
                        /mol_type="mRNA"
                        /cultiyar="Bright Yellow No.2"
                        /db_xref="taxon:4097"
                        /clone="BY9044"
                        /cell_line="BY-2"
                        /note="Vector: pGEM-T easy; primer: M13 forward; mRNA
                        obtained from lag, lag and stationary phase cells"
ORIGIN
Query Match      87.6%; Score 18.4; DB 5; Length 668;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      2 CCGATGATGATTACTCTCC 21
        |||||||
        298 CCGATGATGATTACTCTCC 317

Db

RESULT 3
LOCUS      CF439390                      764 bp    mRNA    linear    EST 04-SEP-2003
DEFINITION EST675735 normalized cDNA library of onion Allium cepa cDNA clone
ACCESSION  CF439390
VERSION     CF439390.1 GI:34462080
KEYWORDS   EST.
SOURCE      Allium cepa (onion)
ORGANISM   Allium cepa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
            Allium.
REFERENCE  1 (bases 1 to 764)
            Havey, M.J., Cheung, F., Van Aken, S., Uteback, T. and Town, C.D.
            Expressed Sequence Tags from a normalized library of mixed onion
            tissues (Allium cepa)
            Unpublished (2003)
JOURNAL    Contact: Havey MW
            Department of Horticulture
            USDA-ARS and University of Wisconsin
            1575 Linden Drive, Madison, WI 53706, USA
            Tel: 608-262-1830
            Fax: 608-262-4743
            Email: mjhavey@facstaff.wisc.edu
COMMENT

```

```

TIGR sequence name ACAAX777R. For more information:
http://haveylab.hort.wisc.edu
Seq primer: CAG GAA ACA GCT ATG ACC.
Location/Qualifiers
1..764
/organism="Allium cepa"
/mol_type="mRNA"
/cultivar="Red Creole(bulbs), unknown(callus), Ebano &
Texas Legend(roots)"
/db_xref="taxon:4679"
/clone="ACAAX77"
/tissue_type="Callus, roots, and young bulbs"
/clone_lib="normalized cDNA library of onion"
/note="Vector: pCMVSPORT6.1-ccdb (Invitrogen); Site 1:
EcoRV (5'); Site 2: NotI (3'); Equal molar amounts of mRNA
from callus, roots, and young bulbs were combined to
synthesize the library. Normalization to enrich for
low-copy transcripts was performed by proprietary
techniques of Invitrogen."
ORIGIN
Query Match      85.7%; Score 18; DB 7; Length 764;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      4 GGATGATGATTACTCTCC 21
        |||||||
        622 GGATGATGATTACTCTCC 639

Db

RESULT 4
LOCUS      CL965683                      948 bp    DNA    linear    GSS 21-SEP-2004
DEFINITION OsIFCC012507 Oryza sativa Expressed Library Oryza sativa (indica
cultiyar-group) genomic, genomic survey sequence.
ACCESSION  CL965683
VERSION     CL965683.1 GI:52386049
KEYWORDS   GSS.
SOURCE      Oryza sativa (indica cultivar-group)
ORGANISM   Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE  1 (bases 1 to 948)
            Ma, L., Mang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
            Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
            Wong, G. K. S., Deng, X. W. and Wang, J.
            An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
            Unpublished (2004)
JOURNAL    Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES   Location/Qualifiers
            source          1..948
                        /organism="Oryza sativa (indica cultivar-group)"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:39946"
                        /clone_lib="Oryza sativa Express Library"
                        /note="Oryza sativa exon trapped genomic sequences "
ORIGIN
Query Match      84.8%; Score 17.8; DB 9; Length 948;
Best Local Similarity 90.5%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 TCCGATGATGATTACTCTCC 21

```

```

DB      462 TCCGGATGACATTTCACCACC 482
|||||
RESULT 5 BM472190          983 bp mRNA linear EST 05-FEB-2002
LOCUS    BM472190
DEFINITION AGENCOURT_6469590 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5576402
ACCESSION BM472190
VERSION   BM472190.1 GI:18521232
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 983)
AUTHORS  NIH-MGC http://mgs.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          DNA sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: LILAM12328 row: d column: 03
          High quality sequence stop: 653.
FEATURES
     Source
         1..983
         /organism="Homo sapiens"
         /mol_type="RNA"
         /db_xref="taxon:9606"
         /clone="IMAGE:5576402"
         /library_type="embryonal carcinoma, cell line"
         /lab_host="DH10B (phage-resistant)"
         /note="Oligo: testis; Vector: pCMV-SPORT6; Site 1: NotI;
          Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
          Average insert size 2.5 kb. Library enriched for
          full-length clones and constructed by Life Technologies.
          Note: this is a NIH_MGC library."
ORIGIN
Query Match      84.8%; Score 17.8; DB 4; Length 983;
Best Local Similarity 90.5%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 TCCGGATGACATTTCCTCC 21
        |||||
Db       950 TCCGGAGACATTTACTCCTCC 970
        |||||

RESULT 6 AV340408          200 bp mRNA linear EST 11-NOV-1999
LOCUS    AV340408
DEFINITION MUSCULUS CDNA clone 6430514N02 3', mRNA sequence.
ACCESSION AV340408
VERSION   AV340408.1 GI:6380460
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 200)
AUTHORS  Komno,H., Aikawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
          Fukuda,S., Fukuishi,Y., Harai,A., Hayatsu,N., Hirazane,T., Hori,F.,
          Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
          Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,

```

**TITLE**  
**JOURNAL**  
**COMMENT**

Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
 Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,  
 Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y.,  
 Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomioka, N.,  
 Tsumoda, Y., Watanahki, A., Watanabe, S., Yamamura, T., Yasunishi, A.,  
 Yokoca, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESs (Kono, H., et al. 1999)  
 Unpublished (1999)  
 Contact: Yoshinide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-2 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Sasaki, N., Izawa, M., Watanahki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,  
 Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and  
 Hayashizaki, Y.  
 Transcriptional sequencing: A method for DNA sequencing using RNA  
 polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)  
 Itoh, M., Kitenai, T., Akiyama, J., Shibata, K., Iwama, M., Kawai, J.,  
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
 Okazaki, Y. and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. *Genome Res.* 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,  
 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
 further details.

**Location/Qualifiers**  
 1..200

```

/oranism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="6430514N02"
/sex="male"
/tissue_type="olfactory brain"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male
olfactory bulb"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAAGACATCCAGACCTCTTTTCTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adaptor of
sequence [5' GAGAGAGAGATTCTCGAGTTAAATAATTAATCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified Bluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI"
```

RESULT 7  
LOCUS CO809020  
DEFINITION AGNCOURT\_30341687 NIH\_ZGC\_5 Danilo rerio cDNA clone IMAGE:7396239  
5', mRNA sequence.  
ACCESSION CO809020  
VERSION CO809020.1 GI:51026281  
KEYWORDS EST.  
SOURCE Danilo rerio (zebrafish)  
ORGANISM Danilo rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Osteichthyes;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 805)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
JOURNAL Contact: Daniela S. Gerhard, Ph.D.  
COMMENT Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgsaps-remail.nih.gov  
Tissue Procurement: John Postlethwait, University of Oregon  
cDNA Library Preparation: Dr. Sumio Sugano  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: L1M615579 row: 1 column: 13  
High quality sequence stop: 501.  
Location/Qualifiers  
1..805  
/organism="Danilo rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:7396239"  
/tissue\_type="contains eggs from one female, from all  
stages of development as well as support cells"  
/lab\_host="DH10B Tona"  
/clone\_lib="NIH\_ZGC\_5"  
/note="Organ: ovary; Vector: pME18-F13; Site:1; DraIII;  
Site:2: DraIII; 1st strand cDNA was primed with an  
oligo(dT) primer  
[GGGCTGACAGAGCGCTATGCGCTTTTCTTTTCTTTT];  
double-stranded cDNA was ligated to a DraIII adaptor  
[GGCTUACUG], digested and directionally cloned into  
distinct DraIII sites of the pME18-F13. Library was size  
selected for 1.0 kb, with a average insert size of ~1.2kb.  
Library constructed by Yutaka Suzuki (University of Tokyo  
Institute of Medical Science). Custom primers recommended  
for sequencing: 5' end primer 5'-GGATGTCCTTACTTCTCA-3',  
and 3' end primer 5'-GCACTCGACTCGACACA-3'. Note: This  
is a Zebrafish Gene Collection (ZGC) library"

ORIGIN  
Query Match 82.9%; Score 17.4; DB 7; Length 805;  
Best Local Similarity 94.7%; Pred. No. 5.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 TCCGGATGATGATTACT 19  
Db 750 TCCGGAAGATGATTACT 768

RESULT 8  
LOCUS CA328214  
DEFINITION UI-M-FY0-cdb-k-07-0-UI.r1 NIH BMAP\_FY0 Mus musculus cDNA clone  
IMAGE:6827360 5', mRNA sequence.  
ACCESSION CA328214  
VERSION CA328214.1 GI:24546312  
KEYWORDS EST.

SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgsaps-remail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: PYX-5.  
Location/Qualifiers  
1..196  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6827360"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 13.5, 14.5, 16.5, 17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_FY0"  
/note="Organ: Brain; Vector: PYX-Asc; Site:1: EcoR I;  
Site:2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into PYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is AGCAGACAG. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
program coordinator."

ORIGIN  
Query Match 81.0%; Score 17; DB 6; Length 196;  
Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 3 CGGATGATGATTACT 19  
Db 86 CGGATGATGATTACT 102

RESULT 9  
LOCUS AA254497  
DEFINITION val7c04.r1 Soares mouse lymph node BMDM1 Mus musculus cDNA clone  
IMAGE:723174 5', mRNA sequence.  
ACCESSION AA254497  
VERSION AA254497.1 GI:1889091  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,





/db\_xref="taxon:10090"  
/clone="NIA:10900H05 IMAGE:30000088"  
/issue\_type="Newborn Kidney"  
/dev\_stage="Newborn"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse Newborn Kidney cDNA library (Long)"  
/note="Organ: Kidney; Vector: pSPORT1 (Invitrogen);  
Site 1: Sali; Site 2: NotI; Mouse cDNA project by the  
Laboratory of Genetics, National Institute on Aging (NIA),  
Intramural Research Program, NIH  
(http://lgen.grc.nia.nih.gov/cDNA). This is a  
long-transcript enriched cDNA library (Ref. Genome Res.  
11:1553-1558 (2001). [PMID:11544191]). In  
brief, double-stranded cDNAs were synthesized with an  
oligo(dT) primer (Invitrogen: 5'-  
pGACTAGTTCAGATCGGAGCGCCGCTTTTCTTTT-3') from 26  
microgram of total RNA, treated with T4 DNA polymerase,  
and purified by ethanol-precipitation. The cDNAs were  
ligated to lone-linker UL-Sali4, purified by  
phenol/chloroform, and separated from free linkers by  
centricion 100. Then, the cDNAs were amplified by  
long-range high fidelity PCR using Ex Tag polymerase  
(Takara) with a primer Sali-S. The products were purified  
by phenol/chloroform and centricion 100. The cDNAs were  
digested with Sali and NotI enzymes, and cloned into  
Sali/NotI site of pSPORT1 plasmid vector. The DH10B E.  
coli host was transformed with ligation mixture by the  
standard chemical method. The average insert size is about  
3.0 kb. The library was constructed by Yulan Piao (NIA)."

## ORIGIN

Query Match 81.0%; Score 17; DB 4; Length 320;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 CGGATGATGATTACCT 19  
|||  
Db 92 CGGATGATGATTACCT 76

RESULT 12  
BE634382/c 326 bp mRNA linear EST 25-AUG-2000  
LOCUS uv67b02.v1 Soares mouse 3bDMS Mus musculus cDNA clone IMAGE:3412203  
DEFINITION BE634382  
5', mRNA sequence.  
ACCESSION BE634382.1 GI:9917070  
VERSION EST.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 326)  
REFERENCE NCI-CCAF http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapb-remail.nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:1088015  
Seq primer: -40RP from Gtlbco  
High quality sequence stop: 316.  
Location/Qualifiers

## FEATURES

1..326  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3412203"  
/sex="male"  
/issue\_type="Spleen"

## ORIGIN

Query Match 81.0%; Score 17; DB 2; Length 326;  
Best Local Similarity 100.0%; Pred. No. 7.7e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 CGGATGATGATTACCT 19  
|||  
Db 86 CGGATGATGATTACCT 70

RESULT 13  
BX525225 337 bp mRNA linear EST 27-JUN-2003  
LOCUS BX525225  
DEFINITION BX525225 Soares mouse lymph node NbMLN Mus musculus cDNA clone  
IMAGE:5210532 ; IMAGE:723174, mRNA sequence.  
ACCESSION BX525225.1 GI:32317508  
VERSION EST.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 337)  
REFERENCE Hell, O., Ebert, L., Neubert, P., Peters, M., Radelof, U., Schneider, D.  
AUTHORS and Korn, B.  
TITLE Mouse Unigeneset - RZPD2  
JOURNAL Unpublished (2003)  
COMMENT Contact: Ina Rolfe  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
RZPD; IMGP95210532.  
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;  
Mouse Unigeneset - RZPD2 (RZPDLIB No.981)  
http://www.rzpd.de/Cloncards/cgi-bin/showlib.pl.cgi/response?libNo=981 Contact: Ina Rolfe  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de  
This clone is available royalty-free from RZPD,  
contact RZPD (clone@rzpd.de) for further information. Seq primer:  
T7, Primer sequence: TAATGCACTCACTATAGG.

## FEATURES

1..337  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5210532 ; IMAGE:723174"  
/sex="male"  
/issue\_type="lymph node"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse lymph node NbMLN"  
/note="Organ: lymph node; Vector: pRT73D-Pac (Pharmacia)  
with a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5'



Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submision Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Tissues were provided by Michela Fagioli and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
 source  
 location/Qualifiers  
 1. 401  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="K230310G10"  
 /tissue\_type="visual cortex"  
 /clone\_lib="RIKEN full-length enriched, visual cortex"

# ORIGIN

Query Match 81.0%; Score 17; DB 6; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 7.9e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGGATGATGATTTACTT 19  
 |||||  
 Db 304 CGGATGATGATTTACTT 320

Search completed: November 7, 2005, 09:18:30  
 Job time : 2109.89 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:11:16 ; Search time 668.064 Seconds  
(without alignments)  
1233.024 Million cell updates/sec

Title: US-10-777-131a-2

Perfect score: 17  
Sequence: 1 tccgccggcggtcgggac 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 2422767955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_ets:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	100.0	17	6	AX770104
2	17	100.0	823	6	AX770121
3	17	100.0	3002	6	CQ880370
4	17	100.0	40056	1	MSGY348
5	17	100.0	110000	1	AE000516_12
6	17	100.0	327650	1	BX248337
7	17	100.0	349306	1	BX842575
8	16	94.1	1566	1	MGU95316
9	16	94.1	134943	2	AC148866
10	16	94.1	300704	1	AE017316
11	16	94.1	314100	1	SC0939106
12	15.4	90.6	356	10	AF133913
13	15.4	90.6	795	3	PVU491803
14	15.4	90.6	1034	8	RATPGP
15	15.4	90.6	1109	8	AK111058
16	15.4	90.6	1194	6	A48328
17	15.4	90.6	1194	6	ARI98358
18	15.4	90.6	1415	6	CQ777515
19	15.4	90.6	1415	10	MMU13705

20	15.4	90.6	1442	10	BC049235	BC049235 Mus muscu
21	15.4	90.6	1448	10	BC003339	BC003339 Mus muscu
22	15.4	90.6	1478	10	BC062227	BC062227 Rattus no
23	15.4	90.6	1498	10	BC061950	BC061950 Mus muscu
24	15.4	90.6	1528	6	BD203690	BD203690 Human nuc
25	15.4	90.6	1528	6	AX014845	AX014845 Sequence
26	15.4	90.6	1652	6	BD203774	BD203774 Human nuc
27	15.4	90.6	1652	6	AX015057	AX015057 Sequence
28	15.4	90.6	1762	10	BC019550	BC019550 Mus muscu
29	15.4	90.6	1775	1	AF224501	AF224501 Streptomy
30	15.4	90.6	2637	4	AB112589	AB112589 Fells cat
31	15.4	90.6	4496	6	A48326	A48326 Sequence 4
32	15.4	90.6	4496	6	ARI98356	ARI98356 Sequence
33	15.4	90.6	5945	1	AY426768	AY426768 Streptomy
34	15.4	90.6	5985	1	SCU43704	SCU43704 Streptomy
35	15.4	90.6	6375	1	AY632768	AY632768 Mycobacte
36	15.4	90.6	8988	1	MTBKATA	MTBKATA Mycobacte
37	15.4	90.6	15065	1	AE005094	AE005094 Halobacte
38	15.4	90.6	70575	2	AC025103	AC025103 Homo sapi
39	15.4	90.6	84985	1	SNA278573	SNA278573 Streptomy
40	15.4	90.6	110000	1	AE016822_01	AE016822_01 Continuation (2 of
41	15.4	90.6	110000	2	AE016822_02	AE016822_02 Continuation (3 of
42	15.4	90.6	110000	2	AC106698_2	AC106698_2 Continuation (7 of
43	15.4	90.6	110000	2	LMFLCHR15_6	LMFLCHR15_6 Tetradon
44	15.4	90.6	129768	2	AC134955	AC134955 Tetradon
45	15.4	90.6	138809	2	AC151220	AC151220 Bos tauru

#### ALIGNMENTS

RESULT 1	AX770104	17 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770104	Sequence 2 from Patent WO03016562.			
DEFINITION	AX770104				
ACCESSION	AX770104				
VERSION	AX770104.1	GI:32437682			
KEYWORDS					
SOURCE					
ORGANISM		synthetic construct			
REFERENCE		1			
AUTHORS		Gicquel, B.			
TITLE		Compositions and methods for detecting multidrug resistant strains of M. tuberculosis having mutations in genes of the mtrA family			
JOURNAL		Patent: WO 03016562-A 2 27-FEB-2003;			
FEATURES		INSTITUT PASTEUR (FR)			
source		location/Qualifiers			
		1..17			
		/organism="synthetic construct"			
		/mol_type="unassigned DNA"			
		/db_xref="taxon:32630"			
		/note="Primer"			
ORIGIN					
Query Match		100.0%; Score 17; DB 6; Length 17;			
Best Local Similarity		100.0%; Pred. No. 5.5e+03;			
Matches		17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy		1 TCCGCCGGGTGCGGAC 17			
Db		1 TCCGCCGGGTGCGGAC 17			
RESULT 2					
LOCUS	AX770121	823 bp	DNA	linear	PAT 02-JUL-2003
DEFINITION	Sequence 19 from Patent WO03016562.				
ACCESSION	AX770121				
VERSION	AX770121.1	GI:32437699			
KEYWORDS					
SOURCE					
ORGANISM		Mycobacterium tuberculosis			

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE  
1 Gicquel, B.  
Compositions and methods for detecting multidrug resistant strains  
of M. tuberculosis having mutations in genes of the mult family  
Patent: WO 03016562-A 19 27-FEB-2003;  
INSTITUT PASTEUR (FR)

JOURNAL  
location/Qualifiers

FEATURES  
source  
1. 823  
/organism="Mycobacterium tuberculosis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 17; DB 6; Length 823;  
Best Local Similarity 100.0%; Pred. No. 2.5e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  
1 TCCGCCGGGTCCGGGAC 17  
|||||  
Db 820 TCCGCCGGGTCCGGGAC 804

RESULT 3  
CO880370 3002 bp DNA linear PAT 11-OCT-2004  
LOCUS  
DEFINITION Sequence 1 from Patent WO2004083459.  
ACCESSION CO880370  
VERSION CO880370.1 GI:54034136  
KEYWORDS  
SOURCE Mycobacterium tuberculosis  
ORGANISM Mycobacterium tuberculosis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE  
1 Bange, F.C.  
Method and kit for a specific detection of m.tuberculosis  
Patent: WO 2004083459-A 1 30-SEP-2004;  
Arcus Gesellschaft Fuer Molekularbiologische Diagnostik und  
Entwicklung MBH. (DE)

FEATURES  
source  
location/Qualifiers  
1. 3002  
/organism="Mycobacterium tuberculosis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 17; DB 6; Length 3002;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  
1 TCCGCCGGGTCCGGGAC 17  
|||||  
Db 1603 TCCGCCGGGTCCGGGAC 1587

RESULT 4  
MSGY348 40056 bp DNA linear BCT 10-DEC-1996  
LOCUS  
DEFINITION Mycobacterium tuberculosis sequence from clone y348.  
ACCESSION AD000020  
VERSION AD000020.1 GI:11717739  
KEYWORDS  
SOURCE Mycobacterium tuberculosis  
ORGANISM Mycobacterium tuberculosis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE  
1 (bases 1 to 40056)  
Du, L.

TITLE Direct Submission  
JOURNAL Submitted (11-OCT-1996) L. Du, Genome Therapeutics Corporation, 100  
Beaver Street, Waltham, MA, USA, 02154 du@rtic.com  
COMMENT GSDB:S:1004716.  
FEATURES  
source  
location/Qualifiers  
1. 40056  
/organism="Mycobacterium tuberculosis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1773"  
/clone="y348"

ORIGIN

Query Match 100.0%; Score 17; DB 1; Length 40056;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  
1 TCCGCCGGGTCCGGGAC 17  
|||||  
Db 18734 TCCGCCGGGTCCGGGAC 18718

RESULT 5  
AE000516\_12/c  
wPCOMMENT  
Sequence split into 44 fragments LOCUS AE000516 Accession AE000516  
Fragment Name Begin End  
AE000516\_00 1 110000  
AE000516\_01 100001 210000  
AE000516\_02 200001 310000  
AE000516\_03 300001 410000  
AE000516\_04 400001 510000  
AE000516\_05 500001 610000  
AE000516\_06 600001 710000  
AE000516\_07 700001 810000  
AE000516\_08 800001 910000  
AE000516\_09 900001 1010000  
AE000516\_10 1000001 1110000  
AE000516\_11 1100001 1210000  
AE000516\_12 1200001 1310000  
AE000516\_13 1300001 1410000  
AE000516\_14 1400001 1510000  
AE000516\_15 1500001 1610000  
AE000516\_16 1600001 1710000  
AE000516\_17 1700001 1810000  
AE000516\_18 1800001 1910000  
AE000516\_19 1900001 2010000  
AE000516\_20 2000001 2110000  
AE000516\_21 2100001 2210000  
AE000516\_22 2200001 2310000  
AE000516\_23 2300001 2410000  
AE000516\_24 2400001 2510000  
AE000516\_25 2500001 2610000  
AE000516\_26 2600001 2710000  
AE000516\_27 2700001 2810000  
AE000516\_28 2800001 2910000  
AE000516\_29 2900001 3010000  
AE000516\_30 3000001 3110000  
AE000516\_31 3100001 3210000  
AE000516\_32 3200001 3310000  
AE000516\_33 3300001 3410000  
AE000516\_34 3400001 3510000  
AE000516\_35 3500001 3610000  
AE000516\_36 3600001 3710000  
AE000516\_37 3700001 3810000  
AE000516\_38 3800001 3910000  
AE000516\_39 3900001 4010000  
AE000516\_40 4000001 4110000  
AE000516\_41 4100001 4210000  
AE000516\_42 4200001 4310000  
AE000516\_43 4300001 4403837

Continuation (13 of 44) of AE000516 from base 1200001 (AE000516 Mycobacterium tuberculosis  
Query Match 100.0%; Score 17; DB 1; Length 110000;

Best Local Similarity 100.0%; Pred. No. 9.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTCCGGGAC 17  
|||||  
Db 86702 TCCGCCGGGTCCGGGAC 86686

RESULT 6  
BX248337/c 327650 bp DNA linear BCT 06-MAY-2004  
LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment  
DEFINITION 4/14.  
ACCESSION BX248337 BX248333  
VERSION BX248337.1 GI:31617663  
KEYWORDS complete genome.  
SOURCE Mycobacterium bovis AF2122/97  
ORGANISM Mycobacterium bovis AF2122/97  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE 1  
AUTHORS Garnier,T., Eigleier,K., Camus,J.-C., Medina,N., Mansoor,H.,  
Pryor,M., Dutfoy,S., Grondin,S., Lacroix,C., Monsempé,C., Simon,S.,  
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,  
Hewinson,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and  
Hewinson,G.  
The complete genome sequence of Mycobacterium bovis  
Unpublished  
2 (bases 1 to 327650)  
Garnier,T.  
DIRECT SUBMISSION  
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire  
Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex  
15, France. e-mail:cgarnier@pasteur.fr Submitted on behalf of the  
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary  
Laboratories Agency Weybridge, Wobham Lane, New Haw, Addlestone,  
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK. Pr4 Annotation, Genopole, Institut  
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28  
rue du Docteur Roux, 75724 Paris Cedex 15, France  
Location/Qualifiers

FEATURES  
source 1..327650  
/organism="Mycobacterium bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:233413"  
698..2170  
/locus\_tag="Mb0912"  
698..2170  
/locus\_tag="Mb0912"  
/note="Mb0912", -, len: 490 aa. Equivalent to Rv0888, len:  
490 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 490 aa overlap). Probable exported  
protein. Equivalent to AKK45157.1 from Mycobacterium  
tuberculosis strain CDC1551 (507 aa) but shorter 17 aa.  
Contains possible N-terminal signal sequence."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE EXPORTED PROTEIN"  
/protein\_id="CAD93773.1"  
/db\_xref="GI:31617664"  
/db\_xref="GOA:Q10549"  
/db\_xref="UniProt/Swiss-Prot:Q10549"  
/translation="MDYAKRIGQVGLAVVIGVGAATVTHAIGSAAPTPDPSSSTDSP  
VDACPLGGSASLSLAIPGASVPGVQVQVPGSIPDDILNALIDFLAVANGVPII  
ENRTPANQGVSEGGTVGVRDADCPGNRMTFAVRBGAQGHQGVIVDOR  
TASIFYTADPGVGTDPSSVNVSDTSLHVGIGLAPFGHGDVAATVAVNGTPT  
DTISGDSMLTYNAGLPLPSSALILPRFPYTKELIGKRLNYVAVNAVYOHQFLI  
KSKRPSQTPPEPPTLMLPIGVPSDGLNTLSSEFVQRLDRQTYECTSDNCLTKGF  
TYSQRLPGSDQVADVYNLHTNTGGSPPTNANLAQVANYIQNSAGRAVIVGDENARY  
SDDSALLQFAQVNVGLTDAWVQVEHGPTTPPAFTCVGNBCELDLKIIFYSGQGVTL

gene

CDS

gene  
CDS

OAVSGNBAPEKFNKCEPLSDHSPAVVGFHYVADNVAVR"  
complement(4205..3326)  
/gene="cItA"  
/locus\_tag="Mb0913c"  
complement(2205..3326)  
/gene="cItA"  
/locus\_tag="Mb0913c"  
/EC\_number="2.3.3.1"  
/note="Mb0913c, cItA, len: 373 aa. Equivalent to Rv0889c,  
len: 373 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 373 aa overlap). Probable cItA  
(alternate gene name: gItA), citrate synthase 2 (EC  
4.1.3.7), highly similar to others e.g.  
CAB95899.1|AL35998 putative citrate synthase from  
Streptomyces coelicolor (387 aa); P39191|CISY\_BACSU  
citrate synthase II from Bacillus subtilis (366 aa), FASTA  
scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa  
overlap); etc. Also similar to Rv0896|MTCY31.24 from  
Mycobacterium tuberculosis (29.2% identity in 274 aa  
overlap) and Rv1131. Contains P500480 Citrate synthase  
signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE CITRATE SYNTHASE II CITA"  
/protein\_id="CAD93774.1"  
/db\_xref="GI:31617665"  
/db\_xref="GOA:Q10529"  
/db\_xref="UniProt/Swiss-Prot:Q10529"  
/translation="MTVPENFVPGDVGVAFTTEIAEPDQKGLARYGVIEDLV  
QRTVFGVWALVDVGNFSGSLPPAPPEPLPHSDVDVQAGLAMPVIGVAPLD  
IDDAFARQQLARASVMALSYVAGARGIYQVAPQRIIDECSTYARMTWQSEPD  
RHTEIDAIVWYSAEHGMASTFTFARVYASGADVAAALSGAIGSLGHPARV  
LPMIDVEVRAGDASVYKGLIDREKLMGRRHYRVEDPPARVLRDAERLGRVY  
VAVAEQALSELRERRPDRATEVNEFMAVVIDPARVPANMPAMFTCGRTAGWCA  
HLEQKRLKLVRSALYVGGPSPSSVDSEMERVLTGA"  
complement(3413..6061)  
/locus\_tag="Mb0914c"  
complement(3413..6061)  
/locus\_tag="Mb0914c"  
/note="Mb0914c", -, len: 882 aa. Equivalent to Rv0890c,  
len: 882 aa, from Mycobacterium tuberculosis strain H37Rv,  
(99.9% identity in 882 aa overlap). Probable  
transcriptional regulatory protein, luxR family, highly  
similar (but shorter 238 aa in N-terminus) to  
NP\_302202.1|NC\_002677 possible transcriptional regulator  
from Mycobacterium leprae (1106 aa). Also highly similar  
(generally in part) to others e.g. T50568 probable  
multi-domain regulatory protein from Streptomyces  
coelicolor (1334 aa); P10957|NMR1\_ECOLI nitrate/nitrite  
response regulator protein from Escherichia coli (216 aa)  
FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 99  
aa overlap); etc. Also highly similar to others from  
Mycobacterium tuberculosis e.g. MTG02B10.22, MTG008.44,  
MTG036.21, and MTG031.24. Contains P500017 ATP/GTP-binding  
site motif A (P-loop), P500622 Bacterial regulatory  
proteins, luxR family signature, and probable helix-turn  
helix motif from aa 836 to 857 (Score 159, +4.50 SD).  
BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL  
REGULATORS."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN  
(PROBABLY LUXR-FAMILY)"  
/protein\_id="CAD93775.1"  
/db\_xref="GI:31617666"  
/db\_xref="UniProt/Swiss-Prot:P59969"  
/translation="IMRALNQNLVLTGCGVQKRTLAIQIASASELRDGLCEVDLA  
PITSGVVAATAAVAGLPDQPGSTDSLRFPGRMLMLVNCHELDAACALV  
ELDAGCELTILARSIPGAGBITRVSMSITDVALEFAPARASVQGFVOR  
NAAVGEICRILQIPIAIFARVASMELPRLADIDDCFLAGVAGVAFACOT  
LRASIDSHALITTEOILRRLAPFVGGPDLAVRVAAASDDPPSVLDQULTLV  
KSLVADDCQRTRETLFVRRVALKRDLDSGADVAHRRDYYTLAASLNPADN  
DHQRLVARBETEDNLRAAFWRSRENCHITELQLASLPITWGRALHRLGSLWFSN



```

/transf1 table=11
/product="POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN"
/protein_id="CAD93776.1"
/db_xref="GI:3161767"
/db_xref="UniProt/Swiss-Prot:P59970"
/translation="MLFNAVHNSLPNPNDIDHAIIRGSDHPPTCAKCVARGISALGS
LDLRHSLRSCVAPPDVGRCEVPERRRVLIAQGLDVSRLPTGTVTVLLADVDEEST
HLMQCPMEPMATAIHLHDTHYSBAITNMGQVAPVGRGDSFVAAFATSAAPACALCL
LORTSLPFRRLRIGLHTGSVOLRDLYOPTINFRARLRDLAGGOVVLNATGSDLVLT
GRLPADAMVLDGRRHPLRGLPREBWMQLCHDRIDREKRPPRLTAKSSPTSLIPQFTT
FYGRPAQJIS"
gene
7318..8805
/locus_tag="Mb0916"
7318..8805
DSD

```

```

1 TCCGCCGGTCCGGAC 17
|||||
301586 TCCGCCGGTCCGGAC 301570

```

gene  
CDS

```

Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M\_tuberculosis/).
Location/Qualifiers
1. 349306
/organism="Mycobacterium tuberculosis H37Rv"
/mol_type="genomic DNA"
/strain="H37Rv"
/db_xref="taxon:83332"
complement(53. 844)
/locus_tag="RV0927C"
complement(53. 844)
/locus_tag="RV0927C"
/EC_number="1.-.-.-"
/function="UNKNOWN; POSSIBLY INVOLVED IN CELLULAR
METABOLISM."
/note="RV0927C, (MTCY21C12.21C), len: 263 aa. Probable
short-chain dehydrogenase/reductase (EC 1.-.-.-), similar
to various dehydrogenases/reductases, notably
7-alpha-hydroxysteroid dehydrogenases and glucose
1-dehydrogenases e.g. P2539 [HDNA_ECOLI
7-alpha-hydroxysteroid dehydrogenase from Escherichia coli
(255 aa), FASTA scores: opt: 551, E(): 1e-26, (39.5*
identity in 248 aa overlap); NP_252778.1 [NC_002516

```



probable short-chain dehydrogenase from *Pseudomonas aeruginosa* (253 aa); AAC44307.1|US9433.3-keracyl-acyl carrier protein reductase from *Bacillus subtilis* (246 aa); etc. Also similar to other dehydrogenases from *Mycobacterium tuberculosis* e.g. MTCY09P9.36, E():1.4e-18; MTCY369.14, E():8e-17; MTCY02B10.14, E():2.5e-14; MTCY09P9.23c, E():1.5e-13; MTCY03C7.07, E():1.9e-13. Contains PS00061 Short-chain dehydrogenases/reductases family signature, and PS00017 ATP/GTP-binding site motif A (P-loop). BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY."

/codon\_start=1  
/transl\_table=1  
/product="PROBABLE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE"  
/protein\_id="CAB08497.1"  
/db\_xref="GI:2078076"  
/translation="MILDMFLDDKAVITGGGRIAGAIALAFQAQADVLLASRTS SELDAVQIRAGRAHVAADLAHPVTAQLAGVAFGKLDIVNNVGTGMENT LLSTKDLADFAFNVTGAHLLVAAYPLMEHSGGSGVINISITMRARGRAY GTAKAALHYRIALDI.CPRRVNVAIAPSGITSLALVVAANDELRAPEMCATPLRR LGDPVIAAAVYIAPSPAGSFLTGKTLVSDGLTFPNLDLPDL"

/complement(113..136)  
/locus\_tag="Rv0927c"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
/complement(320..406)  
/locus\_tag="Rv0927c"  
/note="PS00061 Short-chain alcohol dehydrogenase family signature"  
/gene="pbcS3"  
/locus\_tag="Rv0928"  
1116..2228  
/locus\_tag="Rv0928"  
/gene="pbcS3"  
/locus\_tag="Rv0928"  
/function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT). THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT."

/standard\_name="phoS2"  
/note="RV0928, (MTCY21C12.22), len: 370 aa. pbcS3 (previously known as phoS2), phosphate-binding lipoprotein component of inorganic phosphate transport system (see citations below), highly similar to others from *Mycobacterium leprae* e.g. Q50099|pbcS3|PHOS1 phosphate-binding protein 3 precursor (328 aa), PASTA scores: opt: 1772, E(): 0, (79.6% identity in 328 aa overlap); and highly similar to others e.g. AAF74819.1|AF137360.1|AF137360 periplasmic phosphate permease from *Mycobacterium avium* (369 aa). Also highly similar to Rv0932c|MTCY08D9.07|pbcS2 PHOSPHATE-BINDING PERIPLASMIC LIPOPROTEIN (370 aa); and Rv0934|pbcS1 PHOSPHATE-BINDING PERIPLASMIC LIPOPROTEIN (374 aa) from *Mycobacterium tuberculosis* (*Mycobacterium tuberculosis* seems to have three pbcS-like proteins, others being Rv0932c and Rv0934c). Contains lipoprotein signature (PS00013) at N-terminus. BELONGS TO FAMILY OF PHOSPHATE RECEPTORS FOR BACTERIAL ABC-TYPE LIPOPROTEIN TRANSPORTERS."

/codon\_start=1  
/evidence=experimental  
/transl\_table=1  
/product="PERIPLASMIC PHOSPHATE-BINDING LIPOPROTEIN PBCS3 (PBP-3) (pbcS3) (PHOS1)"  
/protein\_id="CAE55337.1"  
/db\_xref="GI:41353620"  
/translation="MKLNRFGAAGVLAAGALVLSACGNDNVGAGATGQASAKV CGGKTLKASSTAGANMTRPVNFEOACGGOTLNTYANGSAGISFBNQNDPFG SDVPLKDEAAAARCGSPAMNLPLVNGPLAVTNTASLSINDGTLAKITNGSI TOMNPAIQALNDRFTLPGERLHYVFRSDSGTDDNPNRYIQAAISGAGAGASFG GVGEGANQDGTSLAAKNTPGSTITNTWSRQAOHLTMANTVTSAGGDPAITITDSV GQTLAGATISGAGNDLVLTDSFRRPKRPGSYPIVLAIVETVCSKYPDSQVGTAVKAF LOSTTGAQSGIGDNGYIPIIDEPKRSRLSTAVNMA"

misc\_feature 1152..1184  
/gene="pbcS3"  
/locus\_tag="Rv0928"  
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"  
2241..3215  
/gene="pbcS2"  
/locus\_tag="Rv0929"  
2241..3215  
/gene="pbcS2"  
/locus\_tag="Rv0929"  
/function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT."  
/note="Rv0929, (MTCY21C12.23), len: 324 aa. pbcS2, phosphate-transport integral membrane ABC transporter (see citations below), highly similar to others e.g. NP\_302394.1|NC\_002677 membrane-bound component of phosphate transport from *Mycobacterium leprae* (319 aa); CAB88474.1|ALJ53816 phosphate ABC transporter system permease protein from *Streptomyces coelicolor* (336 aa); NP\_290359.1|NC\_002655 high-affinity phosphate-specific transport system (cytoplasmic membrane component) from *Escherichia coli* strain O157:H7 (319 aa); etc. Also similar to Rv935|MTCY08D9.04|pbcS1 PROBABLE TRANSMEMBRANE ABC TRANSPORTER COMPONENT OF PHOSPHATE UPTAKE SYSTEM from *Mycobacterium tuberculosis* (338 aa). Contains binding-protein-dependent transport systems inner membrane component signature (PS00402)."  
/codon\_start=1  
/transl\_table=1  
/product="PHOSPHATE-TRANSPORT INTEGRAL MEMBRANE ABC TRANSPORTER PBC2"  
/protein\_id="CAB08489.1"  
/db\_xref="GI:3261752"  
/translation="MNTPELTPALVAVDMRPARGRLFKLASAGSTIVAILI AIFLVAVPSLRNHNANFFSTQPTSDSQAEGRDLEMTALSSITLVLAVPV AVGIAVFLTHYAPRRLSRPGAWDLAAVPSIIFGLMGIVLAPKSEIPLFRLNL GMFLFRQGNVSLAGGTTIFPAGIVLSEVMSIISREVFRTPLQIDRALALGA TKWVVMNTVLPYRSGSVVAASMGIRALGDETVAVVILSARPGTWSLPDGYTF ASKIASAASESEPLPGAYISAGFALVFLFLVNAABAIAAGKVNQ"  
2844..2930  
/gene="pbcS2"  
/locus\_tag="Rv0929"  
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp signature"  
3212..4138  
/gene="pbcA1"  
/locus\_tag="Rv0930"  
3212..4138  
/gene="pbcA1"  
/locus\_tag="Rv0930"  
/function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE."

Query Match 100.0%; Score 17; DB 1; Length 349306;  
Best Local Similarity 100.0%; Pred. No. 7.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGCCCGGTGGGGAC 17  
Db 253427 TCCGCCCGGTGGGGAC 253411

RESULT 8  
LOCUS MGU95316 1566 bp DNA linear BCT 11-MAY-1997  
DEFINITION *Mycobacterium gordonae* IS1512 transposase gene, complete cds, and Tn554 tnpb transposase homolog gene, partial cds.  
ACCESSION U95316

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	source
UN95316.1	GI:2078346		Mycobacterium gordonae						
			Mycobacterium gordonae						
			Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.						
			1 (bases 1 to 1566)						
			Picardau, M. and Vincent, V.						
			Identification and characterization of IS-like elements in Mycobacterium gordonae						
			Unpublished						
			2 (bases 1 to 1566)						
			Picardau, M. and Vincent, V.						
			Direct Submission						
			Submitted (27-MAR-1997) laboratoire de Reference des Mycobacteries, Institut Pasteur, 25 rue du Docteur Roux, Paris 75013, France						
			Location/Qualifiers						
			1. 1566						
			/organism="Mycobacterium gordonae"						
			/mol_type="genomic DNA"						
			/strain="960592"						
			/db_xref="taxon:1778"						
			1. >893						
			/organism="Mycobacterium gordonae"						
			/mol_type="genomic DNA"						
			/strain="960592"						
			/db_xref="taxon:1778"						
			/insertion_seq="IS1512"						
			/note="IS256 family member; truncated"						
			54. .893						
			/codon_start=1						
			/transl_table=11						
			/product="IS1512 transposase"						
			/protein_id="AAB54014.1"						
			/db_xref="GI:2078347"						
			/translation="MTAAHDIDPALLAEKLTTHPDVRELLATFHTTLMGASADALCGAAGQSSERTINSRNGTRHRQPTFRAGSLDLAIPIKRGQSIFPDMLERKKAERGLTVATCYLLGVSTRMDKLVETLGTIGLSKQVSVAKELDAVENFRTPLDAGLTFVMAADLVKVRKRAQGVVVAHALIAGVNADQVREILGIDVTVAEDGACMLTFVLSGLTARLSGVALVTSDAHAGLISALIGATLPGASWQRCRTHVATINRPSWLESRRQVGRGVAA"						
			/complement(858. >1566)						
			/note="similar to Tns54 tnpb transposase"						
			/codon_start=2						
			/transl_table=11						
			/product="transposase homolog"						
			/protein_id="AAB54015.1"						
			/db_xref="GI:2078348"						
			/translation="WITRGGRORALPHRRGRILTFVHEGCRPTSRLLKGLIDAAVAAARRGSDKLIHTPHQLRHTRETSLSINGIGIPALMALMGHTTPMTLYAKLKAATPRAVYVANDKRVAGLLPLTAVANAPLPIDKAWLAEMLKTRLAHGCAAPQAPAGPCPRANICEQDFVPDPAGLSTINAGLINDIRALQADQARGWHDAAARRQVAASTDOHLQRLRDPSDLPs"						
ORIGIN									
Query Match	94.1%	Score 16;	DB 1;	Length 1566;					
Best Local Similarity	100.0%	Pred. No. 6e+03;							
Matches 16;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
QY	2	CCGCCGGGTCCGGGAC	17						
DB	1025	CCGCCGGGTCCGGGAC	1040						
RESULT 9									
AC148866/c	AC148866	134943 bp	DNA	linear	HTG 08-MAY-2004				
LOCUS									
DEFINITION	Felis catus clone RP86-360D22, WORKING DRAFT SEQUENCE, 8 ordered pieces.								
ACCESSION	AC148866								
VERSION	AC148866.2	GI:47084442							
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.								
SOURCE	Felis catus (cat)								

ORGANISM Felis catus

REFERENCE 1 (bases 1 to 134943)

AUTHORS Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B., Coleman,H., Daki,N., Engle,T., Granite,S., Guan,X., Gupta,J., Hadji,I.P., Han,J., Hansen,N., Ho,S.-U., Hu,P., Hulse,B., Iodl,J.R., Jones,C., Karlene,E., Kim,H., Kwong,P., Latic,P., Larson,S., Lee-Lin,S.-Q., Legaspi,R., Maduro,O.I., Maduro,V.B., Margulies,E.H., Masiello,C., Masketi,B., McDowell,J., Mullikin,J.C., Paguirigan,C., Portnoy,M.E., Prasad,A., Puri,O., Reddi-Dugue,N., Schandler,K., Scheller,M.G., Shah,K., Sison,C., Stenricop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Weheby,K.D., Young,A. and Green,E.D.

JOURNAL NISC Comparative Sequencing Initiative

REFERENCE 2 (bases 1 to 134943)

AUTHORS Green,E.D.

JOURNAL Direct Submission

REFERENCE Submitted (14-APR-2004) NIH Intramural Sequencing Center, 8717

AUTHORS Government Circle, Galtherzburg, MD 20877, USA

JOURNAL 3 (bases 1 to 134943)

REFERENCE Green,E.D.

AUTHORS Direct Submission

JOURNAL Submitted (08-MAY-2004) NIH Intramural Sequencing Center, 8717

REFERENCE Government Circle, Galtherzburg, MD 20877, USA

AUTHORS On May 8, 2004 this sequence version replaced gi:46367628.

JOURNAL ----- Genome Center

COMMENT Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc\_zoo@nigrl.nih.gov

----- Project Information

Center project name: efn

Center clone name: 360D22

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 133378 bases at least Q40

Consensus quality: 133780 bases at least Q30

Consensus quality: 134067 bases at least Q20

Insert size: 123000; agarose-fp

Insert size: 134243; sum-of-contigs

Quality coverage: 11.0% in Q20 bases; agarose-fp

Quality coverage: 10.17% in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 8 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1

\* 2073 2072: contig of 2072 bp in length

\* 2173 2172: gap of unknown length

\* 2173 6288: contig of 4116 bp in length

*	6389	gap of unknown length
*	6389	44101: contig of 37713 bp in length
*	44102	44201: gap of unknown length
*	44202	66179: contig of 21978 bp in length
*	66180	66279: gap of unknown length
*	66280	68036: contig of 1817 bp in length
*	68037	68186: gap of unknown length
*	68197	72556: contig of 4360 bp in length
*	72557	72655: gap of unknown length
*	72657	113988: contig of 41332 bp in length
*	113988	114088: gap of unknown length
*	114089	contig of 20855 bp in length

```
misc_feature /note="BAC resource: http://bacpac.chori.org/"
1..70717
/note="clone overlaps with GenBank Accession Number
AC092518 clone RP86-16J10 (center project name cfl)
1..2072
/note="assembly_fragment
clone end:SP6
vector_side:left"
2173..6288
/note="assembly_fragment"
6389..44101
/note="assembly_fragment"
44202..66179
/note="assembly_fragment"
66280..68096
/note="assembly_fragment"
68197..72556
/note="assembly_fragment"
72657..113988
/note="assembly_fragment"
107778..1134943
/note="clone overlaps with GenBank Accession Number
AC091696 clone RP86-117J4 (center project name awg)"
114089..134943
/note="assembly_fragment
clone_end:T"
vector_side:right"
```

Query Match	94.1%	Score 16:	DB 2:	length 134943;
Best Local Similarity	100.0%	Pred. No.	2.4e+03;	
Matches	16:	Conservative	0;	Mismatches 0;
		Indels	0;	Gaps 0;
Qy	2	CCCCGGGCTCGGGGAC	17	
Db	5955	CCGCCGGGCTCGGGGAC	5940	

RESULT 10	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AE017316	AE017316	300704 bp DNA linear BCT 27-APR-2004							
		Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough, section 8 of 12 of the complete genome.	AE017316	AE017285					
		AE017316.1 GI:46449852							
		Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough							
		Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough							
		Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobiontales;							
		Desulfobiontaceae; Desulfovibrio.							
		1 (bases 1 to 300704)							
		Heidelberg, J. F., Seebach, R., Hayman, S. A., Hemme, C. L.,							
		Paulsen, I. T., Kolonay, J. F., Eisen, J. A., Ward, N., Methe, B.,							
		Brinkac, L. M., Daugherty, S. C., Deboy, R. T., Dodson, R. J., Durkin, A. S.,							

TITLE	The genome sequence of the anaerobic, sulfate-reducing bacterium <i>Desulfovibrio vulgaris</i> Hildenborough
JOURNAL	Nat. Biotechnol. (2004) In press
PUBMED	15077118
REFERENCE	2 (bases 1 to 300704)
AUTHORS	Heidelberg, J.F., Seshadri, R., Hayeman, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B., Brinkac, L.M., Daugherty, S.C., Desoy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D.E., Haft, D.H., Selengut, J., Peterson, J.D., Davidsen, T.M., Zafar, N., Zhou, L., Ratunec, D., Dimitrov, G., Hance, M., Tan, K., Khouri, H.M., Gill, J., Ureback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and Fraser, C.M.
TITLE	Direct Submission
JOURNAL	Submitted (18-MAR-2004) The Institute for Genomic Research, 9712
FEATURES	Medical Center Dr, Rockville, MD 20850, USA
source	Location/Qualifiers 1..300704

gene  
complement(654. . 881)  
/locus\_tag="DVU2028"  
complement(654. . 881)  
/locus\_tag="DVU2028"  
/note="identified by similarity to ONNI:NT01MC3723"  
/codon\_start=1  
/transl\_table=11  
/product="conserved domain protein"  
/protein\_id="AAS96503.1"  
/db\_xref="GI:46449854"  
/translation="MGATPAPYIRERRRSTEQSLTAQEDGVILELAARSEAFETLSWY  
LSPGHARLISQHLRQIVQELGKARELYR"  
926. .1324  
/locus\_tag="DVU2029"  
926. .1324  
/locus\_tag="DVU2029"  
/locus\_tag="DVU2029"  
/note="identified by match to protein family HMM PF01527"  
/codon\_start=1  
/transl\_table=11  
/product="ISDVu2, transposase OrfA"  
/protein\_id="AAS96504.1"  
/db\_xref="GI:46449855"  
/translation="MSRQSAKEISTVEVVTWQBRWTIAEKLVVESSLPQMSVSR  
VARKGIPNIVFPRKRLMSDGGVYALQADDRVSVABAKALKKRIIDLERLLRKTM"  
EVEIKEDAIIDARRKKLISRSPLPFEDGSL"  
1321. .2163  
/locus\_tag="DVU2030"  
1321. .2163  
/locus\_tag="DVU2030"  
/locus\_tag="DVU2030"  
/note="identified by match to protein family HMM PF00665"

[illegible]

TITLE	Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)
JOURNAL	Nature 417 (6885), 141-147 (2002)
MEDLINE	21996410
PUBMED	12000953
REFERENCE	2 (bases 1 to 314100)
AUTHORS	Bentley S.D.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT	On or before Oct 26, 2002 this sequence version replaced gi:20520920, gi:20520898, gi:20520901, gi:20520903, gi:20520904, gi:20520905, gi:20520906, gi:20520907, gi:20520926, gi:20520909, gi:20520910, gi:20520805, gi:20520706, gi:20520789, gi:20520843.
FEATURES	Location/Qualifiers
source	1..314100
	/organism="Streptomyces coelicolor A3(2)"
	/mol_type="genomic DNA"
	/strain="A3(2)"
	/db_xref="taxon:100226"
	53..56
RBS	67..945
gene	/gene="SC00550"
	/note="synonym: SCF11.30"
	67..945
	/gene="SC00550"
	/note="synonym: SCF11.30"
	67..945
	/gene="SC00550"
	/note="SCF11.30, possible membrane protein, len: 292 aa; similar to TR-Q9ZBQ4 (EMBL:AL034447) Streptomyces coelicolor putative transmembrane protein, 396 aa; fasta scores: opt: 677 z-score: 505.6 E(): 8.6e-21; 43.7% identity in 309 aa overlap. Contains possible hydrophobic membrane spanning region"
	/codon_start=1
	/transl_table=1
	/product="putative membrane protein"
	/protein_id="CAB59607.1"
	/db_xref="GI:6137052"
	/db_xref="GOA:Q9RK60"
	/db_xref="UniProt/TrEMBL:Q9RK60"
	/translation="MGSRRAAPRPSRAAPRGGRGSRGALALGAVTLTGAVGTLLAVGGDDGASDRVTANRADOASNAEAGCAAPLASASASATNAGRPSPASAEKRP TKTASPTSPKASAPASKAPVRAITAGSGDGRSGGSGSGSGSGSGSGSGSGSGSGSGTGAEAQVLAIVNEERPAAGCSPTVANSRTLRADYSDVMASGWSHTGPDG STASRVEAAGYQWSALGENIAGQADAAVMEAMMNSPGHREIINCSKEIGVGHVFCGGGPMWTDDFCGTR"
	complement(1031..2359)
gene	/gene="SC00551"
	/note="synonym: SCF11.31c"
	complement(1031..2359)
	/gene="SC00551"
	/note="SCF11.31c, possible histidine kinase protein, len: 442 aa; similar to TR:O50497 (EMBL:AL020958) Streptomyces coelicolor two-component system sensor, 481 aa; fasta scores: opt: 401 z-score: 451.3 E(): 9e-18; 30.2% identity in 367 aa overlap. Contains Pfam matches to entries PF00512 signal, histidine kinase and PF00672 DUF5, domain found in bacterial signal proteins. Contains possible N-terminal signal peptide sequence"
	/codon_start=1
	/transl_table=1
	/product="putative histidine kinase protein"
	/protein_id="CAB59608.1"
	/db_xref="GI:6137053"
	/db_xref="GOA:Q9RK59"
	/db_xref="UniProt/TrEMBL:Q9RK59"
	/translation="MRRALAGIALAVTSVTLFLVLALLVBEQARDVTAAEORA AALSPVALTTRRADVQQAVALGSAQLVRLPDGDFVGPPAPPEALRAAGRET LADTGAAGMAYLOVTVRGRVAVVEVVAADLTRVATGGMALLAVGIVGSVTLVADTGAAYVRSRGLKRASLALGSGGLGVAVPDGPPEIOWGAAATMYADVDLLAVREIVADLSHRILRTPLTLLYLEADMGATPSALRTKAGQGLEREIDSTITAAKPLAAQVLGTAGTACDVAEVAAMRJDPSVLAQQDREFSLTRPAPVAPVED LAAVVDLLINVEPHRTPOGTRFPVVRVRRRHRLVTDVDAAGPGVADPEALLTGVSVG

```

misc_feature
    GSTGLDITVARAARTADGETTIRAPLGGARVTVSPALDARAAGGDR"
    complement (1070. .1672)
    /gene="SC00051"
    /note="Pfam match to entry PF00512 signal, Histidine
    kinase, score 39.60, E-value 7.1e-08"
    complement (1712. .1921)
    /gene="SC00051"
    /note="Pfam match to entry PF00672 DUF5, Domain found in
    bacterial signal proteins, score 36.90, E-value 4.5e-07"
    complement (2356. .3072)
    /gene="SC00052"
    /note="synonym: SCF11.32c"
    complement (2356. .3072)
    /gene="SC00052"

CDS
    /note="SC00052"

gene
    /note="SCF11.32c, possible response regulator, len: 238
    aa; similar to TR:CA54568 (EMBL:U006392) Streptococcus
    pneumoniae response regulator Rod, 234 aa; faeta scores:
    opt: 583 z-score: 693.9 E(): 2.8e-31; 43.2% identity in
    227 aa overlap. Contains Pfam matches to entries PF00486
    trans reg C, transcriptional regulator protein and
    PF00072 response_reg, response regulator receiver domain"
    /codon_start=1
    /transl_table=1
    /product="putative response regulator"
    /protein_id="CA59609.1"
    /db_xref="GI:6137054"
    /db_xref="GOA:O9RK58"
    /db_xref="UniProt/TrEMBL:O9RK59"
    /translation="MTDRITLCPMASTLVVDDPVITAAALIEVITGCVAKTHQGF
    EALRITSPSPDIVLDLGLPDLGDLVLRITRISVPVLVATARDDEIIRLLNAA
    GADDTWVPSFGQGLAARLAIVLRSPAGIRARPPVLRVADARIDPTARITLADR
    ELPLRRFEDLVAALAAADQVMSRORILAEVQQPYVEDQTVDVHSLRRKQGEVA
    RKPRLTRFRGIGIKLVASAP"
    complement (2380. .2595)
    /gene="SC00052"

misc_feature
    /note="Pfam match to entry PF00486 trans_reg_C,
    transcriptional regulatory protein, C terminal, score
    93.00, E-value 1.7e-26"
    complement (2710. .3042)
    /gene="SC00052"

misc_feature
    /note="Pfam match to entry PF00072 response_reg, Response
    regulator receiver domain, score 123.00, E-value 5.6e-33"
    complement (3090. .3093)
    /gene="SC00053"
    /note="synonym: SCF11.33c"
    complement (3099. .3536)
    /gene="SC00053"
    /note="SCF11.33c, unknown, len: 145 aa"
    /codon_start=1
    /transl_table=1
    /product="hypothetical protein SCF11.33c"
    /protein_id="CA59610.1"
    /db_xref="GI:6137055"
    /db_xref="UniProt/TrEMBL:O9RK57"
    /translation="WKPEPFGTAIRHTPPSPRNTVALPSPSEGVITTPAPSSFGAPAA
    PAGGAGVAVRGLGVGVTAASGSVALVRDGGGLAVRGLGVAIVPLDVGK
    GAGSGFVAVCTGAIVGRVDPPTKWIIVMTVITLAAVQDSQIR"
    complement (3691. .5388)
    /gene="SC00054"
    /note="synonym: manA, SCF11.34c"
    complement (3691. .5388)

CDS
    /note="SCF11.34c, manA, probable secreted
    beta-mannosidase, len: 327 aa; similar to many e.g.
    MANA_STRL1_P51529 mannan endo-1,4-beta-mannosidase
    precursor (EC 3.2.1.78) (363 aa), faeta scores: opt: 1189
    z-score: 1332.7 E(): 0.611% identity in 288 aa overlap.
    Contains N-terminal signal sequence, and pfam match to
    entry PF00150 cellulase, Cellulase (glycosyl) hydrolase
    family 5). Also similar to SC2H4.16 (62.2% identity in 288
    aa overlap)
    SCF11.34c, possible sugar hydrolase, partial CDS, len:

```

>278 aa; C-terminus similar to SM-GUN4 THEFU (EMBL:U20093)  
Thermomonospora fusca endoglucanase E-4 precursor (EC  
3.2.1.4) (endo-1,4-beta-glucanase E-4), CelD, 880 aa;  
fasta scores: opt: 544 z-score: 555.0 E(): 1.5e-23; 42.8%  
identity in 215 aa overlap. Contains pfam matches to  
entriees PF00553 CBD 2, cellulose binding domain and  
PF00041 fn3, fibronectin type III domain. Overlaps with  
Streptomyces coelicolor str73.01c, leni. >272 aa;  
C-terminus similar to many e.g. MANA, str121, ps1529 mannan  
endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363  
aa), fasta scores: opt: 1189 z-score: 1332.7 E(): 0, 61.1%  
identity in 288 aa overlap. Contains N-terminal signal  
sequence, and pfam match to entry PF00150 cellulase,  
cellulase (glycosyl hydrolase family 5)<sup>9</sup>  
/codon\_start=1  
/transl\_table=11  
/product="putative secreted beta-mannosidase"  
/protein\_id="CAD55266.1"  
/db\_xref="GI:24413729"  
/db\_xref="GOA:O8CK53"  
/db\_xref="UniProt/TREMBL:O8CK53"  
/translation="MRPARRPTPTTLPRLITGLAALIGLVVAGALCGAALASPA  
RSAGVPSKGAQAGLHIGDGRLENGNDPVMRGVNAHTYPGSTSLADVKAAGAN  
SVRVSLDGRHRSNGPADVAVALI EOCKANRLICVLEVHDTTGYAEDAAAGTLHDAD  
TWIGLKDVLACQEDYVIVINIGNEPMPGTEPTVAAYVKUAAAGLQHTIVDAP  
NMGDMDGVMAANASVYDADPTGNLIFSIHMYSVPTAOSITLYLNAFVAEPIL  
GEGGPDADVDGDDPDPTMMATABOLRGLYLAWSGNDPULDIADLPBRLSGMGE  
RIRHGVNGIACTSEBATTYGGADPDDEAPAPAPASTANTDSYTLGVAATDDTA  
VSGDVVAVGDSITVAASSTTTATVGLTAATVTAFAVARAAGRSKRSATVEY  
TTDGGTTPGAGSVGVRVGMFGPGFQELAVRNTGSAIGPMTLAFAPADQGVTKM  
MGCTATPDAGAVSAVAPASVATIPAGVTITIPARKDQNTTAPAAFRNLGAVCATP"  
complement(3703..4005)  
/gene="SC00554"

## misc\_feature

Query Match 94.1%; Score 16; DB 1; Length 314100;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 2 CCGCGGGGTCCGGGAC 17  
Db 82543 CCGCGGGGTCCGGGAC 82558

RESULT 12  
AF133913 356 bp mRNA linear ROD 04-NOV-1999  
DEFINITION Mus musculus ARL-6 interacting protein-6 (Aip-6) mRNA, partial cds.  
ACCESSION AF133913  
VERSION AF133913.1 GI:4927207  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Ingley,E., Williams,J.H., Walker,C.E., Teal,S., Colley,S.,  
Sayer,M.S., Tibbrook,P.A., Sarna,M., Beaumont,J.G. and Klinken,S.P.  
TITLE A novel ADP-ribosylation like factor (ARL-6), interacts with the  
protein-conducting channel SSG1beta subunit  
JOURNAL FEBS Lett. 459 (1), 69-74 (1999)  
MEDLINE 99439712  
PUBMED 10508919  
REFERENCE 2 (bases 1 to 356)  
AUTHORS Ingley,E., Teal,S. and Klinken,S.P.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAR-1999) Biochemistry, University of Western  
Australia, Rear 50 Murray St., Perth, WA 6000, Australia  
FEATURES  
location/Qualifiers  
1..356  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone\_lib="yeast two-hybrid EML C.1"

gene <1..>356  
/gene="Aip-6"  
CDS <1..>356  
/gene="Aip-6"  
/codon\_start=1  
/product="ARL-6 interacting protein-6"  
/protein\_id="AA033051.1"  
/db\_xref="GI:4927208"  
/translation="LRRLARRRQVTPAPATAPQVADYTGDDAKQSGDQDNRGCDQV  
ARDRRAPRASGETRRLPRVGGSPVLPRKRGIFPATKATQARRWRIQL  
SLICSLFAVLAEFL"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 10; Length 356;  
Best Local Similarity 94.1%; Pred. No. 1.5e+04;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TCCGCGGGGTCCGGGAC 17  
Db 197 TCCGCGGGGTCCGGGAC 213

RESULT 13  
PVU491803 795 bp mRNA linear INV 08-SEP-2002  
LOCUS Patella vulgata partial mRNA for wnt2 protein.  
DEFINITION Patella vulgata partial mRNA for wnt2 protein.  
ACCESSION AJ491803  
VERSION AJ491803.1 GI:21685397  
KEYWORDS wnt2 protein.  
SOURCE Patella vulgata (common limpet)  
ORGANISM Eukaryota; Metazoa; Mollusca; Gastropoda; Bogastrropoda; Docoglossa;  
Patellina; Patellogoda; Patellicidae; Patella.

REFERENCE  
AUTHORS Prud'homme,B., Lartillot,N., Balavoine,G., Adoutte,A. and  
Vervoort,M.  
TITLE Phylogenetic analysis of the wnt gene family. Insights from  
1ophotrochozoan members  
JOURNAL Curr. Biol. 12 (16), 1395 (2002)  
MEDLINE 22182518  
PUBMED 12194820  
REFERENCE 2 (bases 1 to 795)  
AUTHORS Vervoort,M.  
TITLE Direct Submission  
JOURNAL Submitted (26-JUN-2002) Vervoort M., Evolution et Developpement,  
Centre de Genetique Moleculaire, 1 Av. de la Terrasse, 91198  
Gif-sur-Yvette, FRANCE  
FEATURES  
source  
1..795  
location/Qualifiers  
/organism="Patella vulgata"  
/mol\_type="mRNA"  
/db\_xref="taxon:6465"  
1..795  
/gene="wnt2"  
1..>795  
/gene="wnt2"  
/function="signaling molecule"  
/codon\_start=1  
/product="wnt2 protein"  
/protein\_id="CAD57171.1"  
/db\_xref="GI:21685398"  
/db\_xref="GOA:O8MP14"  
/db\_xref="UniProt/TREMBL:O8MP14"  
/translation="ARGBAAPVSISSAGVHAITRACSKGDLHCACDPTKRGSKD  
SRGFDMGCCSDNVRFGSHFVRMVPVDLKERVRARALMHNHNRARBRARPRKRL  
CKRGVSGSCTVRCWLMQEFRRVGDYTKTKGATQIMINQAGTGIYANRRHKRP  
TRSDLVYFEISPDVICIDIGISGTGVRCKKSLGTDGCDIMCCGGRGFTQTERKL  
QCCCKRHWCTLVCKKCSGLTFIPVRELFHRLSLHLRLRKX"

ORIGIN  
Query Match 90.6%; Score 15.4; DB 3; Length 795;  
Best Local Similarity 94.1%; Pred. No. 1.3e+04;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCTCCGGGTCGGGAC 17  
 DB 371 TTCCTCCGGGTCGGGAC 387

RESULT 14  
 RATPGP 1034 bp mRNA linear ROD 20-JAN-2000  
 LOCUS Rattus norvegicus mRNA for plasma glutathione peroxidase precursor,  
 DEFINITION complete cds.  
 ACCESSION D00680  
 VERSION D00680.1 GI:220843  
 KEYWORDS plasma glutathione peroxidase precursor; plasma glutathione  
 peroxidase; seleno-protein.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 1034)  
 AUTHORS Yoshimura,S., Watanabe,K., Suenizu,H., Onozawa,T., Mizoguchi,J.,  
 Tsuda,K., Hata,H. and Moriuchi,T.  
 TITLE Tissue specific expression of the plasma glutathione peroxidase  
 gene in rat kidney  
 JOURNAL J. Biochem. 109 (6), 918-923 (1991)  
 MEDLINE 92041746  
 PUBMED 1939013  
 COMMENT These data kindly submitted in computer readable form by: Shinichi  
 Yoshimura  
 Dept. Cell. Biol.  
 Tokai University  
 School of Medicine  
 Bohseidai  
 Isehara  
 Kanagawa 259-11  
 Japan  
 Phone: 0463-93-1121 x2583  
 Codon at position 350 to 352 is translated as selenocysteine.  
 FEATURES  
 source Location/Qualifiers  
 1. 1034  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Wistar"  
 /db\_xref="taxon:10116"  
 /clone="pRGPx-3"  
 /tissue\_type="placenta"  
 134..814  
 /EC\_number="1.11.1.9"  
 /note="selenocysteine"  
 /codon\_start=1  
 /transl\_except=(pos:350..352,aa:Sec)  
 /product="plasma glutathione peroxidase precursor"  
 /protein\_id="BA00587.2"  
 /db\_xref="GI:6723180"  
 /translation="MSRTLRASCLSLLAGVPVPRGQSEKTKDCHGMSGTYEVG  
 ALTIDGVEYIKQYAKYILPVNASTYUGITDYEALNADDEKPGCLVITGPCN  
 OPGKQEPENSEILPSLKYRPGGFVNFQFEGDVNGEGCKFTFLKNSCPFA  
 ELUGSPGLFWEPMKIHDIRNFKELVPGDIPIMRWYHRTVSNVMILSYMRQ  
 AALGARGR"  
 134..190  
 191..811  
 /product="plasma glutathione peroxidase"  
 ORIGIN  
 sig\_peptide  
 mat\_peptide  
 Query Match 90.6%; Score 15.4; DB 10; Length 1034;  
 Best Local Similarity 94.1%; Pzed. No. 1.2e+04;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 15  
 AK111058  
 LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:002-175-D04, full  
 DEFINITION insert sequence.  
 ACCESSION AK111058.1 GI:32996267  
 VERSION AK111058  
 KEYWORDS FLI\_CDNA; oligo capping.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1  
 AUTHORS The Rice Full-Length cDNA Consortium, National Institute of  
 Agrobiological Sciences Rice Full-Length cDNA Project Team;  
 Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,  
 Kihimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,  
 Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Ii,C.,  
 Ohtsuki,K., Shishiki,T., Foundation of Advancement of International  
 Science Genome Sequencing & Analysis Group, Otsu,Y., Murakami,K.,  
 Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,  
 Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,  
 Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Nishikura,J.,  
 Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,  
 Kuemegi,T., Oka,M., Ryu,R., Ueda,W., Matsubara,K., RIKEN,  
 Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,  
 Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,  
 Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Otsu,N., Ota,Y.,  
 Saito,R., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shitaki,T.,  
 Yoshino,M. and Hayashizaki,Y.  
 TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from  
 japonica rice  
 JOURNAL Science 301 (5631), 376-379 (2003)  
 MEDLINE 12752273  
 PUBMED 12869764  
 REFERENCE 2 (bases 1 to 1109)  
 AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,  
 Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,  
 Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoaka,T.,  
 Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,  
 Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,  
 Kawagawa,S., Kato,H., Kawagashira,N., Kawai,J., Kawamata,M.,  
 Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,  
 Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
 Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,  
 Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,  
 Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,  
 Nanki,T., Narikawa,R., Nishikura,J., Nishi,K., Nomura,K.,  
 Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,  
 Otsu,N., Ota,Y., Otsu,Y., Ryu,R., Satoh,K., Sakai,C., Sakai,K.,  
 Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,  
 Shingawa,A., Shitaki,T., Shishiki,T., Sogabe,Y., Sugano,S.,  
 Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,  
 Tagawa,A., Takahashi,F., Takaku-Ikai,S., Tanaka,T., Tomaru,A.,  
 Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,  
 Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and  
 Yoshimura,A.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of  
 Agrobiological Sciences, Department of Molecular Genetics, Head of  
 Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki  
 305-8602, Japan (E-mail:skkukuhienia.affrc.go.jp,  
 Tel:81-29-838-7007, Fax:81-29-838-7007)  
 COMMENT This clone is one of the 28k full-length cDNA clones from japonica  
 rice.  
 URL : <http://cdna01.dna.affrc.go.jp/cDNA/>  
 NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K.,  
 Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,  
 Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Nanki,T.,  
 Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and  
 Yamamoto,M.



FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Maeda, H., Mura, U., Mizuno, K., Naitaka, R., Nishikawa, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Teunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.  
Genome Exploration Research Group in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takanashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyota, T., Waki, K., Yaeunishi, A. and Hayashizaki, Y.

## FEATURES

## SOURCE

1. 1109  
location/Qualifiers  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultiyar="Nippondare"  
/db\_xref="taxon:39947"  
/clone="002-175-D04"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 8; Length 1109;  
Best Local Similarity 94.1%; Pred. No. 1.2e+04;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCGGGGAC 17  
|||  
Db 101 TCCGCCGGGTCGGGGAC 117

Search completed: November 7, 2005, 06:41:49  
Job time : 676.064 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21 ; Search time 196.043 Seconds  
(without alignments)  
513.336 Million cell updates/sec

Title: US-10-777-131a-2

Perfect score: 17

Sequence: 1 tccgcgggtcggggac 17

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

1: Geneseqn1980s:\*\n2: Geneseqn1990s:\*\n3: Geneseqn2000s:\*\n4: Geneseqn2001as:\*\n5: Geneseqn2001bs:\*\n6: Geneseqn2002as:\*\n7: Geneseqn2002bs:\*\n8: Geneseqn2003as:\*\n9: Geneseqn2003bs:\*\n10: Geneseqn2003cs:\*\n11: Geneseqn2003ds:\*\n12: Geneseqn2004as:\*\n13: Geneseqn2004bs:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	100.0	17	8	ABZ79928 Mycobacte
2	17	100.0	823	8	ABZ79951 Mycobacte
3	17	100.0	3002	13	ADST75303 Mycobacte
4	17	100.0	110000	4	AA199682_12 Continuation (13 o
5	17	100.0	110000	4	AA199683_12 Continuation (13 o
6	15.4	90.6	205	10	ADDD34800 Mouse mit
7	15.4	90.6	337	12	ADH00311 Kidney di
8	15.4	90.6	707	3	AAA90423 Human pla
9	15.4	90.6	879	8	ACA23119 Prokaryot
10	15.4	90.6	1034	12	ADP72472 Renal tox
11	15.4	90.6	1194	2	AAT58555 Streptom
12	15.4	90.6	1368	8	ACA25382 Prokaryot
13	15.4	90.6	1415	12	ADJ75949 Marker ge
14	15.4	90.6	1528	2	AA277483 Human ova
15	15.4	90.6	4496	2	AAT58553 Streptom
16	15.4	88.2	144	6	ABQ90649 M. capsul
17	15.4	88.2	1167	3	AAC55858 S. lavend
18	15.4	88.2	1167	10	ADE10267 S. lavend
19	15.4	88.2	1362	6	ABQ90990 M. capsul
20	15.4	88.2	2969	6	ABZ11506 Human pol

C	21	15	88.2	2969	12	ADMA4024 Novel hum
	22	15	88.2	3765	3	AAC55782 Nucleotid
	23	15	88.2	3765	10	ADE10194 S. lavend
C	24	15	88.2	30156	12	ADK71085 Human MP2
C	25	15	88.2	53500	3	AAC55842 Complexe
C	26	15	88.2	53500	10	ADE10261 S. lavend
	27	15	88.2	110000	4	AA199682_12 Continuation (13 o
	28	15	88.2	110000	4	AA199682_13 Continuation (13 o
	29	15	88.2	110000	4	AA199683_12 Continuation (14 o
	30	15	88.2	110000	4	AA199683_13 Continuation (14 o
	31	14.4	84.7	250	4	AA136528 Human mus
	32	14.4	84.7	250	8	ABX59516 CDNA enc
	33	14.4	84.7	250	12	ADJ30266 Human mus
	34	14.4	84.7	367	9	ACT13330 DNA clone
	35	14.4	84.7	392	9	ACT13323 DNA clone
C	36	14.4	84.7	464	9	ACH17641 Human adu
	37	14.4	84.7	475	9	ACT13258 DNA clone
	38	14.4	84.7	499	9	ACT13288 DNA clone
	39	14.4	84.7	501	9	ACT13255 DNA clone
	40	14.4	84.7	525	9	ACT13279 DNA clone
	41	14.4	84.7	540	9	ACT13232 DNA clone
	42	14.4	84.7	540	9	ACT13217 DNA clone
	43	14.4	84.7	545	9	ACT13348 DNA clone
	44	14.4	84.7	549	9	ACT13239 DNA clone
	45	14.4	84.7	558	9	ACT13227 DNA clone

## ALIGNMENTS

RESULT 1	ABZ79928	standard; DNA; 17 BP.
ID	ABZ79928	
XX	ABZ79928;	
DT	19-MAY-2003	(first entry)
XX		
DE	Mycobacterium tuberculosis mutT2-2 PCR primer SEQ ID NO:2.	
XX		
KW	Mycobacterium tuberculosis; mutT2; alkA; ogt; Rv3908; mutY; Rv3909;	
KM	detection; multidrug resistance; multiple drug resistance; MDR;	
XX	infection; PCR primer; ss.	
OS	Mycobacterium tuberculosis.	
XX	Synthetic.	
PN	WO2003016562-A2.	
XX		
PD	27-FEB-2003.	
XX		
FP	14-AUG-2002; 2002WO-EP009679.	
XX		
PR	14-AUG-2001; 2001US-0311824P.	
PR	21-AUG-2001; 2001US-0313523P.	
PA	(INSP ) INST PASTEUR.	
XX		
PI	Glcquel B;	
XX		
DR	WPI, 2003-256711/25.	
XX		
PT	Predicting the epidemic character of a Mycobacterium tuberculosis isolate	
PT	and/or the acquisition of multiple drug resistance (MDR) by the isolate	
PT	by detecting an alteration in the DNA repair system of the isolate.	
XX		
PS	Claim 32; Page 16; 83pp; English.	
XX		
CC	The present invention describes a method for predicting the epidemic	
CC	character of a Mycobacterium tuberculosis isolate and/or a selective	
CC	advantage to be maintained in the host and/or the acquisition of multiple	
CC	drug resistance (MDR) by the isolate comprising detecting an alteration	
CC	in the DNA repair system of the isolate. Also described: (1) detecting a	

CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a PCR primer for M. tuberculosis mutT2,  
CC which is used in the exemplification of the present invention  
CC  
XX Sequence 17 BP; 1 A; 6 C; 8 G; 2 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 17; DB 8; Length 17;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Gy 1 TCCGCCGGGTGCGGAC 17  
Db 1 TCCGCCGGGTGCGGAC 17  
RESULT 2  
ABZ79951/C  
ID ABZ79951 standard; DNA; 823 BP.  
XX  
XX ABZ79951;  
XX  
XX 19-MAY-2003 (first entry)  
XX  
XX Mycobacterium tuberculosis mutT2 nucleotide sequence SEQ ID NO:19.  
XX  
XX Mycobacterium tuberculosis; mutT2; alka; ogt; Rv3908; muY; Rv3909;  
XX detection; multidrug resistance; multiple drug resistance; MDR;  
XX infection; gene; ds.  
XX  
XX Mycobacterium tuberculosis.  
XX  
XX WO2003016562-A2.  
XX  
XX 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-EP009679.  
XX  
XX 14-AUG-2001; 2001US-0311824P.  
XX  
XX 21-AUG-2001; 2001US-0313523P.  
XX  
XX (INSP ) INST PASTEUR.  
XX  
XX Glcquel B;  
XX  
XX WPI; 2003-256711/25.  
XX  
XX Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
XX and/or the acquisition of multiple drug resistance (MDR) by the isolate  
XX by detecting an alteration in the DNA repair system of the isolate.  
XX  
XX Disclosure; Fig 4A; 83pp; English.  
XX  
XX The present invention describes a method for predicting the epidemic  
XX character of a Mycobacterium tuberculosis isolate and/or a selective  
XX advantage to be maintained in the host and/or the acquisition of multiple  
XX drug resistance (MDR) by the isolate comprising detecting an alteration  
XX in the DNA repair system of the isolate. Also described: (1) detecting a  
XX Mycobacterium tuberculosis strain with a MDR phenotype; (2) a detecting a  
XX polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
XX an Escherichia coli strain containing the plasmid pMYC2501; and (5)  
XX detecting in a patient infected by Mycobacterium tuberculosis a higher  
XX risk of being unable to eliminate the bacillus or of developing MDR  
XX tuberculosis. The method is useful for predicting the epidemic character  
XX of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
XX be maintained in the host and/or the acquisition of MDR by the isolate.  
XX The present sequence represents a M. tuberculosis mutT2 nucleotide

CC sequence, which is used in the exemplification of the present invention  
XX  
XX Sequence 823 BP; 144 A; 253 C; 277 G; 149 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 17; DB 8; Length 823;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Gy 1 TCCGCCGGGTGCGGAC 17  
Db 820 TCCGCCGGGTGCGGAC 804  
RESULT 3  
ADS75303/C  
ID ADS75303 standard; DNA; 3002 BP.  
XX  
XX ADS75303;  
XX  
XX 16-DEC-2004 (first entry)  
XX  
XX Mycobacterium tuberculosis nargHJI nitrate reductase operon Segid 1.  
XX  
XX nargHJI nitrate reductase operon; single nucleotide polymorphism; SNP;  
XX detection method; ds.  
XX  
XX Mycobacterium tuberculosis.  
XX  
XX  
XX Key Location/Qualifiers  
XX variation replace(1500,C)  
XX /tag= a  
XX /standard\_name= "Single nucleotide polymorphism"  
XX misc\_feature 1717..1719  
XX /tag= b  
XX /note= "GTC start codon"  
XX  
XX WO2004083459-A1.  
XX  
XX 30-SEP-2004.  
XX  
XX 19-MAR-2004; 2004WO-EP002911.  
XX  
XX 20-MAR-2003; 2003DE-01013791.  
XX  
XX (ARTU-) ARTUS GBS MOLBKLARBIOLGISCHE DIAGNOSTI.  
XX  
XX Bange F;  
XX  
XX WPI; 2004-728487/71.  
XX  
XX Specific detection of Mycobacterium tuberculosis, useful for differential  
XX diagnosis of infection, by detecting specific polymorphism in the  
XX promoter of the nitrate reductase operon.  
XX  
XX Claim 1; SEQ ID NO 1; 46pp; German.  
XX  
XX This invention relates to a novel method for the detection of  
XX Mycobacterium tuberculosis (M. tuberculosis) occurring in a biological  
XX sample. Specifically, it refers to the amplification of the nargHJI  
XX nitrate reductase operon, including position -215 in the 5' to 3' reading  
XX frame that occurs upstream of the GTC start codon of the operon. The  
XX present invention describes detecting an M. tuberculosis specific  
XX polymorphism at position -215. The polymorphism (C-215T) is detected by  
XX hybridisation to specific probes such that M. tuberculosis can be  
XX differentiated from M. bovis, M. bovis BCG, M. africanum and M. microti  
XX by analysis of the melting curve. The method is used to detect M.  
XX tuberculosis infection, especially to differentiate it from other species  
XX of the M. tuberculosis complex, and furthermore provides rapid and  
XX specific detection, and differentiation of M. tuberculosis. This  
XX polymnucleotide sequence is the M. tuberculosis nargHJI nitrate reductase  
XX operon containing the -215 SNP at position 1500 in this sequence of the  
XX invention.

SQ Sequence 3002 BP; 483 A; 923 C; 1013 G; 583 T; 0 U; 0 Other;

Query Match 100.0%; Score 17; DB 13; Length 3002;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTCCGGGAC 17  
DB 1603 TCCGCCGGGTCCGGGAC 1587

## RESULT 4

AA199682\_12/c

Continuation (13 of 45) of AA199682 from base 1200001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682

WP	Fragment Name	Begin	End
WP	AA199682_00	1	110000
WP	AA199682_01	100001	210000
WP	AA199682_02	200001	310000
WP	AA199682_03	300001	410000
WP	AA199682_04	400001	510000
WP	AA199682_05	500001	610000
WP	AA199682_06	600001	710000
WP	AA199682_07	700001	810000
WP	AA199682_08	800001	910000
WP	AA199682_09	900001	1010000
WP	AA199682_10	1000001	1110000
WP	AA199682_11	1100001	1210000
WP	AA199682_12	1200001	1310000
WP	AA199682_13	1300001	1410000
WP	AA199682_14	1400001	1510000
WP	AA199682_15	1500001	1610000
WP	AA199682_16	1600001	1710000
WP	AA199682_17	1700001	1810000
WP	AA199682_18	1800001	1910000
WP	AA199682_19	1900001	2010000
WP	AA199682_20	2000001	2110000
WP	AA199682_21	2100001	2210000
WP	AA199682_22	2200001	2310000
WP	AA199682_23	2300001	2410000
WP	AA199682_24	2400001	2510000
WP	AA199682_25	2500001	2610000
WP	AA199682_26	2600001	2710000
WP	AA199682_27	2700001	2810000
WP	AA199682_28	2800001	2910000
WP	AA199682_29	2900001	3010000
WP	AA199682_30	3000001	3110000
WP	AA199682_31	3100001	3210000
WP	AA199682_32	3200001	3310000
WP	AA199682_33	3300001	3410000
WP	AA199682_34	3400001	3510000
WP	AA199682_35	3500001	3610000
WP	AA199682_36	3600001	3710000
WP	AA199682_37	3700001	3810000
WP	AA199682_38	3800001	3910000
WP	AA199682_39	3900001	4010000
WP	AA199682_40	4000001	4110000
WP	AA199682_41	4100001	4210000
WP	AA199682_42	4200001	4310000
WP	AA199682_43	4300001	4410000
WP	AA199682_44	4400001	4411529

Query Match 100.0%; Score 17; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTCCGGGAC 17  
DB 87212 TCCGCCGGGTCCGGGAC 87196

## RESULT 5

AA199683\_12/c

Continuation (13 of 44) of AA199683 from base 1200001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683

WP	Fragment Name	Begin	End
WP	AA199683_00	1	110000
WP	AA199683_01	100001	210000
WP	AA199683_02	200001	310000
WP	AA199683_03	300001	410000
WP	AA199683_04	400001	510000
WP	AA199683_05	500001	610000
WP	AA199683_06	600001	710000
WP	AA199683_07	700001	810000
WP	AA199683_08	800001	910000
WP	AA199683_09	900001	1010000
WP	AA199683_10	1000001	1110000
WP	AA199683_11	1100001	1210000
WP	AA199683_12	1200001	1310000
WP	AA199683_13	1300001	1410000
WP	AA199683_14	1400001	1510000
WP	AA199683_15	1500001	1610000
WP	AA199683_16	1600001	1710000
WP	AA199683_17	1700001	1810000
WP	AA199683_18	1800001	1910000
WP	AA199683_19	1900001	2010000
WP	AA199683_20	2000001	2110000
WP	AA199683_21	2100001	2210000
WP	AA199683_22	2200001	2310000
WP	AA199683_23	2300001	2410000
WP	AA199683_24	2400001	2510000
WP	AA199683_25	2500001	2610000
WP	AA199683_26	2600001	2710000
WP	AA199683_27	2700001	2810000
WP	AA199683_28	2800001	2910000
WP	AA199683_29	2900001	3010000
WP	AA199683_30	3000001	3110000
WP	AA199683_31	3100001	3210000
WP	AA199683_32	3200001	3310000
WP	AA199683_33	3300001	3410000
WP	AA199683_34	3400001	3510000
WP	AA199683_35	3500001	3610000
WP	AA199683_36	3600001	3710000
WP	AA199683_37	3700001	3810000
WP	AA199683_38	3800001	3910000
WP	AA199683_39	3900001	4010000
WP	AA199683_40	4000001	4110000
WP	AA199683_41	4100001	4210000
WP	AA199683_42	4200001	4310000
WP	AA199683_43	4300001	4403765

Query Match 100.0%; Score 17; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTCCGGGAC 17  
DB 86681 TCCGCCGGGTCCGGGAC 86665

## RESULT 6

ADD34800

ID ADD34800 standard; DNA; 205 BP.

AC ADD34800;

DT 15-JAN-2004 (first entry)

DE Mouse mitochondrial DNA sequence SEQ ID NO:2579.

KW ds; mouse; array; mitochondrial; hybridisation; energy-metabolism;

KW mitochondrial disease; oxidative phosphorylation dysfunction;

KW oxidative stress; apoptosis; aging.

OS Mus musculus.

XX

```
PN WO2003020220-A2.
XX
XX 13-MAR-2003.
PD
XX
XX 30-AUG-2002; 2002WO-US027886.
PF
XX 30-AUG-2001; 2001US-0316323P.
XX 31-AUG-2001; 2001CA-02356540.
XX
XX (UYEM-) UNIV EMORY.
PA
XX Wallace DC, Levy S, Kerstam K, Procaccio V;
PI WPI; 2003-300821/29.
XX
XX Array containing probes for genes involved in mitochondrial biology,
PT useful for determining mitochondrial biology gene expression profiles for
PT use in diagnosing pathologies and identifying biochemical pathways.
XX
XX Claim 2; SEQ ID NO 2579; 201pp; English.
XX
XX The invention relates to a novel array comprising at least two isolated
XX nucleotide molecules, each molecule having a sequence capable of uniquely
XX hybridizing to a nucleic acid molecule which is an expression product of
XX a gene involved in mitochondrial biology. The array comprises two or more
XX isolated nucleic acid molecules or spots, each molecule having a sequence
XX chosen from sequence of 994 human probes and 2046 mouse probes. An array
XX of the invention is useful for determining an expression profile of a
XX mouse or human sample containing nucleic acid, by contacting the array
XX with the sample under conditions allowing selective hybridization, and
XX measuring hybridization of nucleic acid in the sample to the array to
XX produce an expression profile. The array is also useful for determining
XX an expression profile of a first labelled sample containing nucleic acid
XX relative to a second, differently labelled sample containing nucleic
XX acid. The second sample is a reference or a standard. An array is useful
XX for determining an expression profile diagnostic of an energy-metabolism-
XX related physiological condition. An array of the invention is useful for
XX determining mitochondrial biology gene expression profiles of organisms,
XX such as human, mice and closely related species, tissue and organs of
XX such organisms, which are useful for determining expression profiles
XX diagnostic of energy metabolism-related physiological conditions,
XX diagnosing such physiological conditions, identifying biochemical
XX pathways, genes, and mutations involved in such physiological conditions,
XX identifying therapeutic agents useful for preventing and/or treating such
XX physiological conditions, evaluating and/or monitoring the efficacy of
XX such therapies, and creating and identifying animal models of human
XX energy metabolism-related physiological conditions. An array is also
XX useful for defining expression signatures or profiles for mitochondrial
XX diseases, as well as distinguishing clinical disorders that result from
XX oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
XX apoptosis and aging. An array of the invention contains probes of genes
XX not previously recognised to participate in mitochondrial biology. The
XX sequences shown in ADD33224-ADD35260 represent murine mitochondrial DNA
XX clones used to make the probes of the invention. Some sequences are not
XX present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
XX 1906, 2408 and 2643.
XX
XX Sequence 205 BP; 40 A; 66 C; 58 G; 41 T; 0 U; 0 Other;
SQ
Query Match 90.6%; Score 15.4; DB 10; Length 205;
Best Local Similarity 94.1%; Pred. No. 1.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCCGGGTGCGGAC 17
| | | | | | | | | | | | | | | | | | | | |
DB 85 TCCGCCGGGTGCGGAC 101
| | | | | | | | | | | | | | | | | | | | |
RESULT 7
ADH00311
ID ADH00311 standard; DNA; 337 BP.
XX
XX ADH00311;
```

```
XX
XX 11-MAR-2004 (first entry)
DT
XX
XX Kidney disease-associated gene-related mouse DNA sequence SeqID1440.
DE
XX
XX Kidney disease; proximal tubule; nephrotropic; chronic renal failure;
XX diabetic nephropathy; mouse; murine; ds.
XX
XX Mus sp.
OS
XX
XX WO2003091427-A1.
XX
XX 06-NOV-2003.
PD
XX
XX 23-APR-2003; 2003WO-JP005137.
XX
XX 24-APR-2002; 2002JP-00122253.
XX
XX (KANS-) KANSAI TECHNOLOGY LICENSING ORG CO LTD.
XX
XX Takenaka M, Imai E, Okubo K;
PI WPI; 2004-011769/01.
XX
XX Kidney disease-associated genes expressed in proximal tubule, applicable
XX in diagnosis of, in drug development and in devising appropriate strategy
XX in treatment of e.g. chronic renal failure and diabetic nephropathy.
XX
XX Claim 1; SEQ ID NO 1440; 770pp; Japanese.
XX
XX This invention relates to novel kidney disease-associated genes expressed
XX in proximal tubules. The invention may be useful for the development of
XX compounds with a nephrotropic activity. The genes are applicable in the
XX diagnosis of, in drug development and in devising appropriate strategy in
XX the treatment of (for example) chronic renal failure and diabetic
XX nephropathy. The present sequence is that of a mouse DNA sequence which is
XX related to the invention.
XX
XX Sequence 337 BP; 83 A; 87 C; 85 G; 74 T; 0 U; 8 Other;
SQ
Query Match 90.6%; Score 15.4; DB 12; Length 337;
Best Local Similarity 94.1%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCCGGGTGCGGAC 17
| | | | | | | | | | | | | | | | | | | | |
DB 49 TCCGCCGGGTGCGGAC 65
| | | | | | | | | | | | | | | | | | | | |
RESULT 8
AAA90423
ID AAA90423 standard; CDNA; 707 BP.
XX
XX AAA90423;
AC
XX 10-JAN-2001 (first entry)
DT
XX
XX Human plasma glutathione peroxidase H (GPxH) cDNA.
XX
XX Plasma glutathione peroxidase H; GPxH; human; recombinant production;
XX ss.
XX
XX Homo sapiens.
OS
XX
XX Key
XX CDS
XX
XX Location/Qualifiers
XX
XX 16..696
XX
XX /tag= a
XX /product= "Human GPx4"
XX /transcript= (pos:232..234, aa:Xaa)
XX /note= "These nucleotides represent an in-frame stop
XX codon (TGA); Xaa = unknown"
XX
XX CN1256312-A.
```

```
XX 14-JUN-2000.
PD 29-OCT-1998; 98CN-00121973.
XX
XX 29-OCT-1998; 98CN-00121973.
XX
XX (UYFU-) UNIV FUDAN.
XX
XX Yu L, Tu Q, Fu Q;
XX
XX WPI; 2000-53631/49.
XX
XX P-PSDB; AAB22828.
XX
XX New human glutathione peroxidase and its code sequence, preparation and
XX use.
XX
XX Claim 1; Page 20; 26pp; Chinese.
XX
XX This sequence represents cDNA encoding human plasma glutathione
XX peroxidase H (GPxH). The invention relates to this novel human
XX glutathione peroxidase, nucleic acid encoding it, and to recombinant
XX production of human GPxH. The invention also encompasses applications
XX for human GPxH.
XX
XX Sequence 707 BP; 165 A; 191 C; 192 G; 159 T; 0 U; 0 Other;
SQ
XX
XX Query Match 90.6%; Score 15.4; DB 3; Length 707;
XX Best Local Similarity 94.1%; Pred. No. 1.2e+03;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCCGGGTCGGGAC 17
Db |||||
72 TCCGCCGGGTCGGGAC 88
XX
XX RESULT 9
XX ACA23319/c
XX ID ACA23319 standard; DNA; 879 BP.
XX
XX ACA23319;
XX
XX 19-JUN-2003 (first entry)
XX
XX Prokaryotic essential gene #4976.
XX
XX Antisense; db; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX Borrelia cepacia.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0345923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
XX
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
XX P-PSDB; ABU19449.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
```

```
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 11189; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 879 BP; 178 A; 300 C; 265 G; 136 T; 0 U; 0 Other;
SQ
XX
XX Query Match 90.6%; Score 15.4; DB 8; Length 879;
XX Best Local Similarity 94.1%; Pred. No. 1.1e+03;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCCGGGTCGGGAC 17
Db |||||
53 TCCGCCGGGTCGGGAC 37
XX
XX RESULT 10
XX ADP72472
XX ID ADP72472 standard; DNA; 1034 BP.
XX
XX ADP72472;
XX
XX 26-AUG-2004 (first entry)
XX
XX Renal toxin progression gene marker #1061.
XX
XX db; toxic effect; gene expression profile; kidney tissue;
XX differential gene expression; toxicity progression; toxicity marker;
XX drug screening; toxicity assay; kidney pathology; nephritis;
XX kidney necrosis; glomerular injury; tubular injury;
XX focal segmental glomerulosclerosis.
XX
XX Rattus norvegicus.
XX
XX WO2004048598-A2.
XX
XX 10-JUN-2004.
XX
XX 24-NOV-2003; 2003WO-US037556.
XX
XX 22-NOV-2002; 2002US-00301856.
XX
```

PA (GENE-) GENE LOGIC INC.  
 XX Mendrick DL, Porter MW, Johnson KR, Castle A, Higgs B;  
 PI Elashoff M;  
 XX WPI; 2004-460771/43.  
 DR  
 XX Predicting (the progression of) a toxic effect of a compound, for  
 PT monitoring the progression of renal disease states, comprises preparing a  
 PT gene expression profile of a kidney tissue or cell sample exposed to the  
 PT compound.  
 PS  
 XX Claim 11; SEQ ID NO 1061; 266pp; English.  
 XX  
 XX The invention relates to a method of predicting (the progression of) a  
 CC toxic effect of a compound by preparing a gene expression profile of a  
 CC kidney tissue or cell sample exposed to the compound and comparing the  
 CC gene expression profile to a database, or detecting the level of gene(s)  
 CC expression in a tissue or cell sample exposed to the compound, where  
 CC differential gene expression compared to a control indicates a toxic  
 CC effect (toxicity progression). The method is useful for predicting (the  
 CC progression of) at least one toxic effect of a compound. The genes are  
 CC useful as toxicity markers in drug screening and toxicity assays. The  
 CC methods are useful for predicting the likelihood that a compound or test  
 CC agent will induce various specific kidney pathologies, such as nephritis,  
 CC kidney necrosis, glomerular and tubular injury, or focal segmental  
 CC glomerulosclerosis. The methods are useful for determining the similarity  
 CC of a toxic response to one or more individual compounds and for  
 CC predicting or elucidating the potential cellular pathways influenced,  
 CC induced or modulated by the compound or test agent. The kit is useful for  
 CC predicting or modelling the toxic response of a test compound, for  
 CC monitoring the progression of renal disease states, for identifying genes  
 CC that show promise as new drug targets and for screening known and newly  
 CC designed drugs. This sequence corresponds to a gene marker used in the  
 CC method of the invention. (Note: The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences).  
 XX  
 SQ Sequence 1034 BP; 247 A; 298 C; 277 G; 212 T; 0 U; 0 Other;  
 Query Match 90.6%; Score 15.4; DB 12; Length 1034;  
 Best Local Similarity 94.1%; Pred. No. 1.1e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TCCGCGGGTCGGGAC 17  
 Db 190 TCCGCGGGTCGGGAC 206  
 RESULT 11  
 AAT5855/c  
 ID AAT5855 standard; cDNA; 1194 BP.  
 XX  
 AC AAT5855;  
 XX  
 DT 02-APR-1997 (first entry)  
 XX  
 XX Streptomyces pristinaespiralis snbF gene.  
 DE  
 XX Streptogramin B; antibiotic; biosynthesis; pristinaamycin; virginiamycin;  
 KM pipercolic acid; cyclodeamination; papa; snbA; snbF; p1pA;  
 KM 3-hydroxypicolinic acid; hydroxylation; ds.  
 XX  
 OS Streptomyces pristinaespiralis.  
 XX  
 FH Key Location/Qualifiers  
 FT mat\_peptide 1..1194  
 FT /\*tag= a  
 FT /product= "SnbF"  
 XX  
 PN MO9601901-A1.  
 XX  
 PD 25-JAN-1996.

XX  
 PF 04-JUL-1995; 95WO-FR000889.  
 XX  
 PR 08-JUL-1994; 94FR-00008478.  
 XX  
 XX (RHON ) RHONE POULENC RORER SA.  
 PA  
 PI Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J,  
 PI Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rossat G;  
 XX WPI; 1996-097631/10.  
 DR  
 DR P-PSDB; AAM11585.  
 XX  
 XX New streptogramin B derivative, useful as antibiotics - produced by new  
 PT mutants of Streptomyces having altered genes for streptogramin B  
 PT biosynthesis.  
 PS  
 XX Claim 18; Page 113-114; 146pp; French.  
 XX  
 XX The papa gene of S.pristinaespiralis is involved in the biosynthesis of 4  
 CC -dimethylamino-L-phenylalanine (DMPAPA), a precursor for pristinaamycin  
 CC IA. Upstream of the papa gene, on the complementary strand, is the snbA  
 CC gene coding for 3-hydroxypicolinic acid-AMP ligase. The region between  
 CC these two genes was sequenced and two open reading frames were  
 CC identified. The first (p1pA) decodes to an amino acid sequence with  
 CC homology to ornithine cyclodeaminase from Agrobacterium tumefaciens. The  
 CC p1pA gene product is likely to catalyse the cyclo- deamination of lysine,  
 CC leading to production of pipercolic acid. Mutations in the p1pA gene were  
 CC shown to affect pipercolic acid synthesis but not the synthesis of 3-  
 CC hydroxypicolinic acid. The second open reading frame (snbF) could be  
 CC decoded to give a product with homology to hydroxylases of the cytochrome  
 CC P450 type. Disruption of the p1pA and snbF genes can be used to produce  
 CC strains of S.pristinaespiralis which are unable to produce the antibiotic  
 CC pristinaamycin I but which may be able to produce new, modified forms of  
 CC it  
 XX  
 SQ Sequence 1194 BP; 189 A; 522 C; 321 G; 162 T; 0 U; 0 Other;  
 Query Match 90.6%; Score 15.4; DB 2; Length 1194;  
 Best Local Similarity 94.1%; Pred. No. 1.1e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TCCGCGGGTCGGGAC 17  
 Db 500 TCCGCGGGTCGGGAC 484  
 RESULT 12  
 ACA25382/c  
 ID ACA25382 standard; DNA; 1368 BP.  
 XX  
 AC ACA25382;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 XX Prokaryotic essential gene #7039.  
 DE  
 XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KM drug design; gene.  
 XX  
 OS Burkholderia fungorum.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 XX 21-MAR-2002; 2002WO-US009107.  
 FF  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX P-PSDB; ABU21512.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 13252; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-regulated gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
XX prokaryotic essential genes. Note: The sequence data for this patent did  
XX not form part of the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1368 BP; 156 A; 439 C; 477 G; 296 T; 0 U; 0 Other;  
XX  
XX Query Match 90.6%; Score 15.4; DB 8; Length 1368;  
XX Best Local Similarity 94.1%; Pred. No. 1.1e+03;  
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX 1 TCCGCCGGGTCGGGAC 17  
XX |||||  
XX Db 467 TCCGCCGGGTCGGGAC 451  
XX  
XX  
XX RESULT 13  
XX ADJ75949 standard; DNA; 1415 BP.  
XX  
XX ADJ75949;  
XX  
XX 20-MAY-2004 (first entry)  
XX  
XX Marker gene SEQ ID NO:1201.  
XX  
XX bronchial asthma; chronic obstructive pulmonary disease;  
XX respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;  
XX gene therapy; marker gene; gene; ds.  
XX  
XX Mus musculus.  
XX

XX EPI394274-A2.  
XX  
XX 03-MAR-2004.  
XX  
XX 04-AUG-2003; 2003BP-00254857.  
XX  
XX 06-AUG-2002; 2002JP-00229312.  
XX 20-MAR-2003; 2003JP-00077212.  
XX  
XX (GENO-) GENOX RES INC.  
XX  
XX Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;  
XX WPI; 2004-193155/19.  
XX  
XX Testing for bronchial asthma or chronic obstructive pulmonary disease by  
XX comparing the expression level of a marker gene in a biological sample  
XX from a subject with the expression level of the gene in a sample from a  
XX healthy subject.  
XX  
XX Example 11; SEQ ID NO 1201; 241pp; English.  
XX  
XX The present invention describes a method of testing for bronchial asthma  
XX or chronic obstructive pulmonary disease. The method comprises  
XX determining the expression level of a marker gene in a biological sample  
XX from a subject, comparing the expression level determined with the  
XX expression level of the marker gene in a biological sample from a healthy  
XX subject, and judging whether the subject has bronchial asthma or chronic  
XX obstructive pulmonary disease. The marker gene comprises: (a) a group of  
XX genes (S1) whose expression levels increase when respiratory epithelial  
XX cells are stimulated with interleukin-13; or (b) a group of genes (S2)  
XX whose expression levels decrease when respiratory epithelial cells are  
XX stimulated with interleukin-13. Also described: (1) a reagent (1) for  
XX testing for bronchial asthma or chronic obstructive pulmonary disease;  
XX (2) a kit for screening for a candidate compound for a therapeutic agent  
XX to treat bronchial asthma or chronic obstructive pulmonary disease; (3)  
XX an animal model for bronchial asthma or chronic obstructive pulmonary  
XX disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a  
XX method for producing an animal model for bronchial asthma or chronic  
XX obstructive pulmonary disease; (6) a therapeutic agent for bronchial  
XX asthma or chronic obstructive pulmonary disease, comprising the compound,  
XX a marker gene or an antisense nucleic acid corresponding to a portion of  
XX the marker gene, a ribozyme, a polynucleotide that suppresses the  
XX expression of the gene through an RNAi effect or an antibody recognising  
XX a protein encoded by a marker gene; and (7) a DNA chip for testing for  
XX bronchial asthma or a chronic obstructive pulmonary disease, on which a  
XX probe has been immobilised to assay a marker gene. (1) has respiratory  
XX and antiasthmatic activities, and can be used in gene therapy. The method  
XX is useful for testing for or screening for a therapeutic agent for  
XX bronchial asthma or chronic obstructive pulmonary disease. The present  
XX sequence is used in the exemplification of the present invention.  
XX  
XX Sequence 1415 BP; 319 A; 437 C; 346 G; 313 T; 0 U; 0 Other;  
XX  
XX Query Match 90.6%; Score 15.4; DB 12; Length 1415;  
XX Best Local Similarity 94.1%; Pred. No. 1.1e+03;  
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX 1 TCCGCCGGGTCGGGAC 17  
XX |||||  
XX Db 73 TCCGCCGGGTCGGGAC 89  
XX  
XX  
XX RESULT 14  
XX AA277483 standard; cDNA; 1528 BP.  
XX  
XX AA277483;  
XX  
XX 10-APR-2000 (first entry)  
XX  
XX Human ovarian tumor cDNA library derived EST fragment 34.  
XX  
XX



```

XX XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KM gene therapy; treatment; se.
XX OS Homo sapiens.
XX PN DE19817557-A1.
XX 21-OCT-1999.
XX 09-APR-1998; 98DE-01017557.
XX 09-APR-1998; 98DE-01017557.
XX 09-APR-1998; 98DE-01017557.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX DR WPI; 1999-591920/51.
XX DR P-PSDB; AAY76581, AAY76582.
XX PT New nucleic acid sequences expressed in ovarian, and some other, cancer
XX FT tissues, and derived polypeptides, for treatment of ovarian cancer and
XX FT identification of therapeutic agents.
XX PS Claim 3; Page 165; 310pp; German.
XX CC This invention describes novel nucleic acid (cDNA) sequences (A) which
XX CC have anticancer activity and are highly expressed in ovarian tumor tissue
XX CC (and some also in testis and breast cancer tissue). The products of the
XX CC invention can be used for gene therapy. (A) are used (i) for recombinant
XX CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
XX CC are used (i) to identify agents suitable for treatment of ovarian cancer;
XX CC (ii) directly for treating this form of cancer (including expression from
XX CC gene therapy vectors) and (iii) for generation of specific antibodies.
XX CC (A) are identified by assembling ESTs (expressed sequence tags) from a
XX CC particular tissue type before comparison of expression patterns. This
XX CC allows a significantly longer fragment of the gene to be revealed, so
XX CC should reduce the number of failures associated with the fact that ESTs
XX CC from different libraries may represent different parts of the same
XX CC unknown gene, distorting the estimated frequency of occurrence in a
XX CC particular tissue. AAZ77450-Z77572 represent the human ovarian tumor cDNA
XX CC library derived EST fragments described in the method of the invention
XX CC and encode the protein fragments represented in AAY76505-Y76638
XX SQ Sequence 1528 BP; 365 A; 459 C; 373 G; 331 T; 0 U; 0 Other;
XX
XX Query Match 90.6%; Score 15.4; DB 2; Length 1528;
XX Best Local Similarity 94.1%; Pred. No. 1.1e+03;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCGGGGTCCGGGAC 17
DB 157 TCCGCGGGGTCCGGGAC 173

```

```

FT CDS complement(1..332)
FT /*tag= a
FT /product= "SnbA"
FT /note= "N-terminal coding region only, i.e. a partial
FT open reading frame"
FT CDS
FT /*tag= b
FT /product= "PipA"
FT 1800..2996
FT /*tag= c
FT /product= "SnbF"
FT 3018..4496
FT CDS
FT /*tag= d
FT /product= "Papa"
FT /note= "N-terminal coding region only, i.e. a partial
FT open reading frame"
XX XX MO9601901-A1.
XX PN 25-JAN-1996.
XX PD
XX 04-JUL-1995; 95MO-FR000889.
XX 08-JUL-1994; 94FR-00008478.
XX (RHON ) RHONE POULENC RORER SA.
XX PI Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;
XX PI Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;
XX DR WPI; 1996-097631/10.
XX DR P-PSDB; AAM11584, AAM11585.
XX PT New streptogramin B derivative, useful as antibiotics - produced by new
XX FT mutants of Streptomyces having altered genes for streptogramin B
XX FT biosynthesis.
XX PS Example 1; Page 107-111; 146pp; French.
XX CC The papa gene of S.pristinaespiralis is involved in the biosynthesis of 4
XX CC -dimethylamino-L-phenylalanine (DMPA), a precursor for pristinamycin
XX CC 1A. Upstream of the papa gene, on the complementary strand, is the snbA
XX CC gene coding for 3-hydroxyisovaleric acid-AMP lyase. The region between
XX CC these two genes was sequenced and two open reading frames were
XX CC identified. The first (pipA) decodes to an amino acid sequence with
XX CC homology to ornithine cyclodeaminase from Agrobacterium tumefaciens. The
XX CC pipA gene product is likely to catalyse the cyclo- deamination of lysine,
XX CC leading to production of pipercolic acid. Mutations in the pipA gene were
XX CC shown to affect pipercolic acid synthesis but not the synthesis of 3-
XX CC hydroxyisovaleric acid. The second open reading frame (snbF) could be
XX CC decoded to give a product with homology to hydroxylases of the cytochrome
XX CC P450 type. Disruption of the pipA and snbF genes can be used to produce
XX CC strains of S.pristinaespiralis which are unable to produce the antibiotic
XX CC pristinamycin I but which may be able to produce new, modified forms of
XX SQ Sequence 4496 BP; 659 A; 1980 C; 1321 G; 536 T; 0 U; 0 Other;
XX
XX Query Match 90.6%; Score 15.4; DB 2; Length 4496;
XX Best Local Similarity 94.1%; Pred. No. 9.2e+02;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCGGGGTCCGGGAC 17
DB 2299 TCCGCGGGGTCCGGGAC 2283

```

Search completed: November 7, 2005, 05:40:12  
 Job time : 208.043 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 19:19:37 ; Search time 57.6915 Seconds  
(without alignment)  
482.163 Million cell updates/sec

Title: US-10-777-131A-2  
Perfect score: 17  
Sequence: 1 tcgcgcggcgccggcgac 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 2	17	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 3	15.4	90.6	1194	3 US-08-765-907A-9	Sequence 9, Appli
C 4	15.4	90.6	1194	4 US-09-987-614A-9	Sequence 9, Appli
C 5	15.4	90.6	4496	3 US-08-765-907A-6	Sequence 6, Appli
C 6	15.4	90.6	4496	4 US-09-987-614A-6	Sequence 6, Appli
C 7	15.4	90.6	54986	4 US-09-949-016-16716	Sequence 16716, A
C 8	15	88.2	1167	4 US-09-266-965-97	Sequence 97, Appli
C 9	15	88.2	2969	4 US-09-799-451-388	Sequence 388, App
C 10	15	88.2	3765	4 US-09-266-965-9	Sequence 9, Appli
C 11	15	88.2	53500	4 US-09-266-965-76	Sequence 76, Appli
C 12	15	88.2	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 13	15	88.2	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 14	14.4	84.7	507	4 US-09-902-540-5663	Sequence 5663, Ap
C 15	14.4	84.7	765	4 US-09-252-991A-7684	Sequence 7684, Ap
C 16	14.4	84.7	1032	4 US-09-724-797-75	Sequence 75, Appli
C 17	14.4	84.7	1032	4 US-09-252-991A-9565	Sequence 9565, Ap
C 18	14.4	84.7	1075	2 US-08-809-860-1	Sequence 1, Appli
C 19	14.4	84.7	1281	4 US-09-252-991A-7879	Sequence 7879, Ap
C 20	14.4	84.7	1344	4 US-09-252-991A-9345	Sequence 9345, Ap
C 21	14.4	84.7	1614	4 US-09-252-991A-9523	Sequence 9523, Ap
C 22	14.4	84.7	2112	4 US-09-252-991A-9429	Sequence 9429, Ap
C 23	14.4	84.7	2313	4 US-09-252-991A-7945	Sequence 7945, Ap
C 24	14.4	84.7	2454	4 US-09-252-991A-9316	Sequence 9316, Ap
C 25	14.4	84.7	2647	4 US-09-220-132-77	Sequence 77, Appli
C 26	14.4	84.7	2647	5 PCT-US93-06251-77	Sequence 77, Appli
C 27	14.4	84.7	3173	4 US-09-799-451-428	Sequence 428, App

C 28	14.4	84.7	3390	4 US-09-902-540-6647	Sequence 6647, Ap
C 29	14.4	84.7	3393	4 US-09-902-540-514	Sequence 514, App
C 30	14.4	84.7	3737	4 US-09-747-371-1	Sequence 1, Appli
C 31	14.4	84.7	5074	4 US-09-902-540-691	Sequence 691, App
C 32	14.4	84.7	10892	4 US-09-902-540-962	Sequence 962, App
C 33	14.4	84.7	12237	4 US-09-949-016-17312	Sequence 17312, A
C 34	14.4	84.7	38155	2 US-09-453-702B-79	Sequence 79, Appli
C 35	14.4	84.7	43280	3 US-08-804-237C-1	Sequence 1, Appli
C 36	14.4	84.7	48908	3 US-09-453-702B-137	Sequence 137, App
C 37	14.4	84.7	51354	4 US-09-902-540-1270	Sequence 1270, App
C 38	14.4	84.7	61158	4 US-09-949-016-15041	Sequence 15041, A
C 39	14.4	84.7	75341	4 US-09-949-016-15122	Sequence 15122, A
C 40	14	82.4	340	4 US-09-513-999C-20103	Sequence 20103, A
C 41	14	82.4	465	3 US-09-188-930-49	Sequence 49, Appli
C 42	14	82.4	465	4 US-09-312-283C-49	Sequence 49, Appli
C 43	14	82.4	601	4 US-09-949-016-67201	Sequence 67201, A
C 44	14	82.4	601	4 US-09-949-016-67202	Sequence 67202, A
C 45	14	82.4	833	3 US-09-188-930-247	Sequence 247, App

#### ALIGNMENTS

```
RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      100.0%; Score 17; DB 3; Length 4403765;
Best local similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 TCCGCCGGTCCGGGAC 17
Db      1286681 TCCGCCGGTCCGGGAC 1286665

RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          100.0%; Score 17; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCCGCCGGGTCCGGGAC 17
Db      1287212 TCCGCCGGGTCCGGGAC 1287196

RESULT 3
US-08-765-907A-9/c
; Sequence 9, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-9

Query Match          90.6%; Score 15.4; DB 3; Length 1194;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCCGCCGGGTCCGGGAC 17
Db      500 TCCGCCGGGTCCGGGAC 484

RESULT 4
US-09-987-614A-9/c
; Sequence 9, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
```

```

; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; PRIOR FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-9

Query Match          90.6%; Score 15.4; DB 4; Length 1194;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCCGCCGGGTCCGGGAC 17
Db      500 TCCGCCGGGTCCGGGAC 484

RESULT 5
US-08-765-907A-6/c
; Sequence 6, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-6

Query Match          90.6%; Score 15.4; DB 3; Length 4496;
Best Local Similarity 94.1%; Pred. No. 2.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCCGCCGGGTCCGGGAC 17
Db      2299 TCCGCCGGGTCCGGGAC 2283

RESULT 6
US-09-987-614A-6/c
; Sequence 6, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
```

;; TITLE OF INVENTION: Mutaenthesis  
;; FILE REFERENCE: Streptogramin genes  
;; CURRENT APPLICATION NUMBER: US/09/987,614A  
;; CURRENT FILING DATE: 2001-11-15  
;; PRIOR APPLICATION NUMBER: US/08/765,907  
;; PRIOR FILING DATE: 1997-03-20  
;; NUMBER OF SEQ ID NOS: 17  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 6  
;; LENGTH: 4496  
;; TYPE: DNA  
;; ORGANISM: Streptomyces pristinaespiralis  
US-09-987-614A-6

Query Match 90.6%; Score 15.4; DB 4; Length 4496;  
Best Local Similarity 94.1%; Pred. No. 2.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCGCGGGTCCGGGAC 17  
Db 2299 TCCGCGGGTCCGGGAC 2283

RESULT 7  
US-09-949-016-16716/C  
; Sequence 16716, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16716  
; LENGTH: 54986  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(54986)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-16716

Query Match 90.6%; Score 15.4; DB 4; Length 54986;  
Best Local Similarity 94.1%; Pred. No. 2.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCGCGGGTCCGGGAC 17  
Db 2310 TCCGCGGGTCCGGGAC 2294

RESULT 8  
US-09-266-965-97  
; Sequence 97, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1

;; CURRENT APPLICATION NUMBER: US/09/266,965  
;; CURRENT FILING DATE: 1999-03-12  
;; EARLIER APPLICATION NUMBER: US 08/624,447  
;; EARLIER FILING DATE: 1996-08-19  
;; EARLIER APPLICATION NUMBER: PCT/US94/11279  
;; EARLIER FILING DATE: 1994-10-06  
;; EARLIER APPLICATION NUMBER: US 08/133,963  
;; EARLIER FILING DATE: 1993-10-07  
;; NUMBER OF SEQ ID NOS: 145  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 97  
;; LENGTH: 1167  
;; TYPE: DNA  
;; ORGANISM: Streptomyces lavendulae  
US-09-266-965-97

Query Match 88.2%; Score 15; DB 4; Length 1167;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCGCGGGTCCGGGA 16  
Db 519 CCGCGGGTCCGGGA 533

RESULT 9  
US-09-799-451-388/C  
; Sequence 388, Application US/09799451  
; Patent No. 6783969  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Zhang, Jie  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Ma, Yungting  
; APPLICANT: Yamazaki, Victoria  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Weinman, Tom  
; APPLICANT: Ghosh, Reena  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6783969e1 Nucleic Acids and  
; FILE REFERENCE: Polypeptides  
; CURRENT APPLICATION NUMBER: US/09/799,451  
; CURRENT FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 948  
; SOFTWARE: pc FL\_genes Version 2.0  
; SEQ ID NO 388  
; LENGTH: 29659  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (567)..(2711)  
US-09-799-451-388

Query Match 88.2%; Score 15; DB 4; Length 2969;  
Best Local Similarity 100.0%; Pred. No. 4.6e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCGCGGGTCCGGGA 16  
Db 2818 CCGCGGGTCCGGGA 2804

RESULT 10  
US-09-266-965-9  
Sequence 9, Application US/09266965  
Patent No. 6495348  
GENERAL INFORMATION:  
APPLICANT: Sherman, D  
APPLICANT: Mao, Y  
APPLICANT: Varoglu, M  
APPLICANT: He, M  
APPLICANT: Sheldon, P  
TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
FILE REFERENCE: 600.456US1  
CURRENT FILING DATE: US/09/266,965  
EARLIER FILING DATE: 1999-03-12  
EARLIER APPLICATION NUMBER: US 08/624,447  
EARLIER FILING DATE: 1996-08-19  
EARLIER APPLICATION NUMBER: PCT/US94/11279  
EARLIER FILING DATE: 1994-10-06  
EARLIER APPLICATION NUMBER: US 08/133,963  
NUMBER OF SEQ ID NOS: 145  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9  
LENGTH: 3765  
TYPE: DNA  
ORGANISM: Streptomyces lavendulae  
US-09-266-965-9

Query Match 88.2%; Score 15; DB 4; Length 3765;  
Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CGCCGGGTGGGGA 16  
|||||  
Db 1097 CGCCGGGTGGGGA 1111

RESULT 11  
US-09-266-965-76/c  
Sequence 76, Application US/09266965  
Patent No. 6495348  
GENERAL INFORMATION:  
APPLICANT: Sherman, D  
APPLICANT: Mao, Y  
APPLICANT: Varoglu, M  
APPLICANT: He, M  
APPLICANT: Sheldon, P  
TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
FILE REFERENCE: 600.456US1  
CURRENT FILING DATE: US/09/266,965  
EARLIER FILING DATE: 1999-03-12  
EARLIER APPLICATION NUMBER: US 08/624,447  
EARLIER FILING DATE: 1996-08-19  
EARLIER APPLICATION NUMBER: PCT/US94/11279  
EARLIER FILING DATE: 1994-10-06  
EARLIER APPLICATION NUMBER: US 08/133,963  
NUMBER OF SEQ ID NOS: 145  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 76  
LENGTH: 53500  
TYPE: DNA  
ORGANISM: Streptomyces lavendulae  
US-09-266-965-76

Query Match 88.2%; Score 15; DB 4; Length 53500;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CGCCGGGTGGGGA 16  
|||||  
Db 20939 CGCCGGGTGGGGA 20925

RESULT 12  
US-09-103-840A-2  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT FILING DATE: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 88.2%; Score 15; DB 3; Length 4403765;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 CGCCGGGTGGGAC 17  
|||||  
Db 1306570 CGCCGGGTGGGAC 1306584

RESULT 13  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT FILING DATE: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1

Query Match 88.2%; Score 15; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 CGCCGGGTGGGAC 17  
|||||  
Db 1307101 CGCCGGGTGGGAC 1307115

RESULT 14  
US-09-902-540-5663  
Sequence 5663, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:

```

; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO: 5663
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-5663

```

```

Query Match      84.7%; Score 14.4; DB 4; Length 507;
Best Local Similarity 93.8%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 TCCGCCGGGTCCGGGA 16
        |||||
Db      293 TCCGCCGGGTCCGGGA 308

```

```

RESULT 15
US-09-252-991A-7684/C
; Sequence 7684, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 7684
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7684

```

```

Query Match      84.7%; Score 14.4; DB 4; Length 765;
Best Local Similarity 93.8%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      2 CCGCCGGGTCCGGGAC 17
        |||||
Db      712 CCGCCGGGTCCGGTAC 697

```

Search completed: November 6, 2005, 23:10:09  
 Job time : 32.6915 secs

***This Page Blank (copy)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 20:56:42 ; Search time 423.915 Seconds  
(without alignments)  
331.640 Million cell updates/sec

Title: US-10-777-131A-2

Perfect score: 17  
Sequence: 1 tccgcgggtcggggac 17

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*
- 20: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*
- 21: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*
- 22: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:\*
- 23: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 24: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 25: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 26: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 27: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 28: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	100.0	17	US-10-216-817-2	Sequence 2, Appli
2	17	100.0	17	US-10-777-131A-2	Sequence 2, Appli
3	17	100.0	823	US-10-216-817-19	Sequence 19, Appli
4	17	100.0	823	US-10-777-131A-19	Sequence 19, Appli
5	15.4	90.6	504	US-10-425-115-163326	Sequence 163326,

C 6	15.4	90.6	879	18	US-10-282-122A-11189	Sequence 11189, A
C 7	15.4	90.6	1368	18	US-10-282-122A-13252	Sequence 13252, A
C 8	15.4	90.6	1415	24	US-10-631-467-1201	Sequence 1201, Ap
C 9	15.4	90.6	1758	16	US-10-156-761-6524	Sequence 6524, Ap
C 10	15.4	90.6	9025608	16	US-10-156-761-1	Sequence 1, Appli
C 11	15.4	90.6	9025608	16	US-10-156-761-1	Sequence 1, Appli
C 12	15.4	88.2	442	21	US-10-425-115-130735	Sequence 130735,
C 13	15.4	88.2	474	14	US-10-027-632-195891	Sequence 195891,
C 14	15.4	88.2	474	18	US-10-027-632-195891	Sequence 195891,
C 15	15.4	88.2	1002	19	US-10-424-348-97	Sequence 97, Appl
C 16	15.4	88.2	1167	16	US-10-267-255-97	Sequence 97, Appl
C 17	15.4	88.2	1518	16	US-10-156-761-5098	Sequence 5098, Ap
C 18	15.4	88.2	1524	20	US-10-437-963-9032	Sequence 9032, Ap
C 19	15.4	88.2	1944	20	US-10-437-963-9032	Sequence 9032, Ap
C 20	15.4	88.2	2969	19	US-10-302-172-388	Sequence 388, Ap
C 21	15.4	88.2	3765	16	US-09-953-348-9	Sequence 9, Appli
C 22	15.4	88.2	3765	16	US-10-267-255-9	Sequence 9, Appli
C 23	15.4	88.2	53500	10	US-09-953-348-76	Sequence 76, Appl
C 24	15.4	88.2	53500	10	US-10-267-255-76	Sequence 76, Appl
C 25	15.4	88.2	190	21	US-10-425-115-41023	Sequence 41023, A
C 26	14.4	84.7	250	9	US-09-764-877-2893	Sequence 2893, Ap
C 27	14.4	84.7	250	18	US-10-242-515-2893	Sequence 2893, Ap
C 28	14.4	84.7	310	19	US-10-424-599-76802	Sequence 76802, A
C 29	14.4	84.7	412	20	US-10-437-963-87400	Sequence 87400, A
C 30	14.4	84.7	412	21	US-10-425-115-122422	Sequence 122422,
C 31	14.4	84.7	460	21	US-10-425-115-89259	Sequence 89259, A
C 32	14.4	84.7	464	10	US-09-918-995-8853	Sequence 4853, Ap
C 33	14.4	84.7	509	21	US-10-425-115-19357	Sequence 19357, A
C 34	14.4	84.7	524	19	US-10-424-599-102950	Sequence 102950,
C 35	14.4	84.7	554	20	US-10-437-963-87275	Sequence 87275, A
C 36	14.4	84.7	561	20	US-10-767-701-21299	Sequence 21299, A
C 37	14.4	84.7	563	20	US-10-424-599-21354	Sequence 21354, A
C 38	14.4	84.7	606	16	US-10-156-761-1573	Sequence 1573, Ap
C 39	14.4	84.7	606	16	US-10-156-761-5910	Sequence 5910, Ap
C 40	14.4	84.7	612	21	US-10-425-115-56398	Sequence 56398, A
C 41	14.4	84.7	612	21	US-10-027-632-285049	Sequence 285049,
C 42	14.4	84.7	633	14	US-10-027-632-285049	Sequence 285049,
C 43	14.4	84.7	633	18	US-10-027-632-285049	Sequence 285049,
C 44	14.4	84.7	636	22	US-10-487-901-6727	Sequence 6727, Ap
C 45	14.4	84.7	636	22	US-10-487-901-6727	Sequence 6727, Ap

#### ALIGNMENTS

RESULT 1

US-10-216-817-2

Sequence 2, Application US/10216817

Publication No. US20030129619A1

GENERAL INFORMATION:

APPLICANT: GICQUEL, BRIGITTE

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG

TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS

FILE REFERENCE: 03495, 0233-00000

CURRENT APPLICATION NUMBER: US/10/216,817

CURRENT FILING DATE: 2002-11-13

PRIOR APPLICATION NUMBER: 60/311,824

PRIOR FILING DATE: 2001-08-14

PRIOR APPLICATION NUMBER: 60/313,523

PRIOR FILING DATE: 2001-08-21

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 17

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURES:

OTHER INFORMATION: Description of Artificial Sequence: Primer

US-10-216-817-2

Query Match

Best Local Similarity 100.0%; Score 17; DB 16; Length 17;

Pred. No. 3.1e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TCCGCCGGGTCCGGGAC 17  
Db 1 TCCGCCGGGTCCGGGAC 17

RESULT 2  
US-10-777-131A-2

; Sequence 19, Application US/10777131A  
; Publication No. US20050026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; TITLE OF INVENTION: IN GENES OF THE mult FAMILY  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT APPLICATION NUMBER: US/10/777.131A  
; PRIOR FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO: 2  
; LENGTH: 17  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer  
US-10-777-131A-2

Query Match 100.0%; Score 17; DB 22; Length 17;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGGAC 17  
Db 1 TCCGCCGGGTCCGGGAC 17

RESULT 3  
US-10-216-817-19/c

; Sequence 19, Application US/10216817  
; Publication No. US20030129619A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; TITLE OF INVENTION: IN GENES OF THE mult FAMILY  
; FILE REFERENCE: 03495.0233-00000  
; CURRENT APPLICATION NUMBER: US/10/216.817  
; PRIOR FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 19  
; LENGTH: 823  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-216-817-19

Query Match 100.0%; Score 17; DB 16; Length 823;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGGAC 17

Db 820 TCCGCCGGGTCCGGGAC 804

RESULT 4  
US-10-777-131A-19/c

; Sequence 19, Application US/10777131A  
; Publication No. US20050026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; TITLE OF INVENTION: IN GENES OF THE mult FAMILY  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT APPLICATION NUMBER: US/10/777.131A  
; PRIOR FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO: 19  
; LENGTH: 823  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-777-131A-19

Query Match 100.0%; Score 17; DB 22; Length 823;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGGAC 17  
Db 820 TCCGCCGGGTCCGGGAC 804

RESULT 5  
US-10-425-115-163326/c

; Sequence 163326, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425.115  
; PRIOR FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO: 163326  
; LENGTH: 504  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_80528C.1  
US-10-425-115-163326

Query Match 90.6%; Score 15.4; DB 21; Length 504;  
Best Local Similarity 94.1%; Pred. No. 5.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGGAC 17  
Db 301 TCCGCCGGGTCCGGGCC 285

RESULT 6  
US-10-282-122A-11189/c



```
; Sequence 11189, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haebelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11189
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-282-122A-11189

Query Match          90.6%; Score 15.4; DB 18; Length 879;
Best Local Similarity 94.1%; Pred. No. 4.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 TCCGCCGGGTGCGGAC 17
    |||||
Db 53 TCCGCCGGGTGCGGAC 37

RESULT 7
US-10-282-122A-13252/c
; Sequence 13252, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haebelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
```

```
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13252
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13252

Query Match          90.6%; Score 15.4; DB 18; Length 1368;
Best Local Similarity 94.1%; Pred. No. 3.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 TCCGCCGGGTGCGGAC 17
    |||||
Db 467 TCCGCCGGGTGCGGAC 451

RESULT 8
US-10-631-467-1201
; Sequence 1201, Application US/10631467
; Publication No. US20050208496A1
; GENERAL INFORMATION:
; APPLICANT: Genex Research Inc.
; TITLE OF INVENTION: Method for testing for bronchreal asthma, or chronic obstructive p
; FILE REFERENCE: 3462.1005-000
; CURRENT APPLICATION NUMBER: US/10/631,467
; PRIOR FILING DATE: 2003-07-31
; PRIOR APPLICATION NUMBER: JP 2003-077212
; PRIOR FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: JP 2002-229312
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 2086
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1201
; LENGTH: 1415
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-631-467-1201

Query Match          90.6%; Score 15.4; DB 24; Length 1415;
Best Local Similarity 94.1%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 TCCGCCGGGTGCGGAC 17
    |||||
Db 73 TCCGCCGGGTGCGGAC 89
```

## RESULT 9

US-10-156-761-6524/c  
; Sequence 6524, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6524  
; LENGTH: 1758  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1758)  
US-10-156-761-6524

Query Match Best Local Similarity 90.6%; Score 15.4; DB 16; Length 1758;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCCGGGAC 17  
DB 1055 TCCGCCGGGTCCGGGAC 1039

## RESULT 10

US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 90.6%; Score 15.4; DB 16; Length 9025608;

Best Local Similarity 94.1%; Pred. No. 17;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCCGGGAC 17  
DB 1815959 TCCGCCGGGTCCGGGAC 1815975

## RESULT 11

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match Best Local Similarity 90.6%; Score 15.4; DB 16; Length 9025608;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCCGGGAC 17  
DB 7838294 TCCGCCGGGTCCGGGAC 7838278

## RESULT 12

US-10-425-115-130735/c  
; Sequence 130735, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: LA ROSE, THOMAS J.  
; APPLICANT: KOVALLIC, DAVID K.  
; APPLICANT: ZHOU, YIHUA  
; APPLICANT: CAO, YONGWEI  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 130735  
; LENGTH: 442  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(442)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_50707C.1

Query Match 90.6%; Score 15.4; DB 16; Length 9025608;

US-10-425-115-130735

Query Match 88.2%; Score 15; DB 21; Length 442;

Best Local Similarity 100.0%; Pred. No. 8.3e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTCCGGG 15

DB 385 TCCGCCGGGTCCGGG 371

RESULT 13

US-10-027-632-195991/c

; Sequence 195991, Application US/10027632

; Publication No. US20020198371A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO: 195991

; LENGTH: 474

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-195991

Query Match 88.2%; Score 15; DB 14; Length 474;

Best Local Similarity 100.0%; Pred. No. 8.1e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Indels 0; Gaps 0;

QY 2 CCGCCGGGTCCGGGA 16

DB 161 CCGCCGGGTCCGGGA 147

RESULT 14

US-10-027-632-195991/c

; Sequence 195991, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO: 195991

; LENGTH: 474

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-195991

Query Match 88.2%; Score 15; DB 18; Length 474;

Best Local Similarity 100.0%; Pred. No. 8.1e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Indels 0; Gaps 0;

QY 2 CCGCCGGGTCCGGGA 16

DB 161 CCGCCGGGTCCGGGA 147

RESULT 15

US-10-424-599-27197

; Sequence 27197, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO: 27197

; LENGTH: 1002

; TYPE: DNA

; ORGANISM: Glycine max

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)...(1002)

; OTHER INFORMATION: unsure at all n locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_12455C.1

US-10-424-599-27197

Query Match 88.2%; Score 15; DB 19; Length 1002;

Best Local Similarity 100.0%; Pred. No. 6.3e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTCCGGG 15

DB 77 TCCGCCGGGTCCGGG 91

Search completed: November 7, 2005, 05:21:08

Job time : 459.915 secs

It's Page Book (1950-51)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 1700.72 Seconds  
(without alignments)  
380.481 Million cell updates/sec

Title: US-10-777-131A-2

Perfect score: 17

Sequence: 1 tccgcgcggcgcgggac 17

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 6847908

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_hic:\*

4: gb\_est3:\*

5: gb\_est4:\*

6: gb\_est5:\*

7: gb\_est6:\*

8: gb\_ges1:\*

9: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	94.1	366	5	BY311757 BY311757
2	16	94.1	438	2	BS776822 MY-20-H-0
3	16	94.1	720	4	BG857100 1024049H0
4	16	94.1	932	5	B0838754 AGENCOURT
5	15.4	90.6	193	1	AI316193 u126b02.Y
6	15.4	90.6	217	5	BY000112 BY000112
7	15.4	90.6	218	2	BS561944 BS561944
8	15.4	90.6	218	5	BY000049 BY000049
9	15.4	90.6	219	9	AY18508 Mus muscu
10	15.4	90.6	222	5	BY006303 BY006303
11	15.4	90.6	228	5	BY006302 BY006302
12	15.4	90.6	234	2	BS561733 BS561733
13	15.4	90.6	236	2	BS561740 BS561740
14	15.4	90.6	243	1	AU096437 AU096437
15	15.4	90.6	244	1	AA277884 VC10C05.Y
16	15.4	90.6	247	5	BY000013 BY000013
17	15.4	90.6	248	2	BS564084 BS564084
18	15.4	90.6	250	5	BY339805 BY339805
19	15.4	90.6	251	5	BY338153 BY338153
20	15.4	90.6	254	5	BY006482 BY006482
21	15.4	90.6	265	2	BS562086 BS562086
22	15.4	90.6	265	2	BS564146 BS564146
23	15.4	90.6	266	2	BS561699 BS561699
24	15.4	90.6	268	2	BS561859 BS561859

25	15.4	90.6	270	2	BS561705 BS561705
26	15.4	90.6	271	2	BS561925 BS561925
27	15.4	90.6	272	2	BS561694 BS561694
28	15.4	90.6	274	2	BS561741 BS561741
29	15.4	90.6	274	2	BS561744 BS561744
30	15.4	90.6	274	2	BS561750 BS561750
31	15.4	90.6	274	2	BS561845 BS561845
32	15.4	90.6	277	2	BS561730 BS561730
33	15.4	90.6	277	2	BS562206 BS562206
34	15.4	90.6	277	2	BS562239 BS562239
35	15.4	90.6	281	5	BY318477 BY318477
36	15.4	90.6	283	5	BY034687 BY034687
37	15.4	90.6	292	2	BS562015 BS562015
38	15.4	90.6	295	5	BY336579 BY336579
39	15.4	90.6	309	6	BY789788 BY789788
40	15.4	90.6	313	5	BY332785 BY332785
41	15.4	90.6	313	8	CC200248 CC200248
42	15.4	90.6	316	5	BY000114 BY000114
43	15.4	90.6	324	6	CD552437 CD552437
44	15.4	90.6	325	5	BY151676 BY151676
45	15.4	90.6	326	5	BY346975 BY346975

#### ALIGNMENTS

RESULT 1	BY311757	366 bp	mRNA	linear	EST 11-DEC-2002
LOCUS	BY311757				
DEFINITION	BY311757 RIKEN full-length enriched, stroma cell Mus musculus cDNA				
ACCESSION	clone 1320023J05 5', mRNA sequence.				
VERSION	BY311757				
KEYWORDS	BY311757.1 GI:26502094				
SOURCE	EST.				
ORGANISM	Mus musculus (house mouse)				
REFERENCE	Mus musculus				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus.				
	1 (bases 1 to 366)				
	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikiado, I., Oseato, N., Saito, R., Suzuki, H., Yamana, T., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., But, C., Hume, D. A., Quackenbush, J., Schriber, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusic, V., Chochina, C., Corzani, L. E., Cousins, S., Dalia, E., Dragani, F. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Gilmont, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Kongave, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pettea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. V., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C. A., Serou, M., Shimada, K., Sulciana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wanless, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyntash-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavoian, M., Zhu, Y., Zimmer, A., Carninci, P., Hayata, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Harada, A., Hashizume, W., Imocani, K., Iishi, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Saeki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E., and Hayashizaki, Y.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
JOURNAL	Nature 420, 563-573 (2002)				
MEDLINE	22354683				
COMMENT	1246851				
	Laboratory for Genome Exploration Research Group, RIKEN Genomic				

Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan.  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/

Altawa, K., Akimura, T., Akawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submision  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN Integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Takashi Ishikawa ( Department of Surgery  
2 Yokohama City University 3-9 Fukuura, Kanazawa-Ku, Yokohama  
236-0004 Japan ) whose assistance we gratefully acknowledge.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

#### FEATURES

##### source

1..366  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="FJ320023J05"  
/cell\_type="stroma cell"  
/clone\_lib="RIKEN full-length enriched, stroma cell"

##### ORIGIN

Query Match 94.1%; Score 16; DB 5; Length 366;

Best Local Similarity 94.1%; Pred. NO. 3.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGGAC 17  
|||||  
Db 98 TCCGCCGGGTCCGGGAC 114

RESULT 2  
BE776822 438 bp mRNA linear EST 20-SEP-2000  
LOCUS MY-20-H-01 Plinfestammy Phytophthora infestans cDNA, mRNA sequence.  
ACCESSION BE776822  
VERSION BE776822.1 GI:10230477  
KEYWORDS EST.  
SOURCE Phytophthora infestans (potato late blight agent)  
ORGANISM Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
Phytophthora.

REFERENCE  
1 (bases 1 to 438)  
Kamoun, S., Hraber, P. T., Sobral, B. W. S., Nuss, D. and Govers, F.  
Initial assessment of gene diversity for the oomycete pathogen  
Phytophthora infestans based on expressed sequences  
Fungal Genet. Biol. 28 (2), 94-106 (1999)  
20056376  
JOURNAL MEDLINE  
10587472  
PUBMED  
Contact: Govers F  
Laboratory of Phytopathology

Wageningen University  
Binnenhaven 9, P.O.Box 8025, 6700 EF, Wageningen, The Netherlands  
Tel: 31 317 483 138  
Fax: 31 317 483 412  
Email: Francine.Govers@medew.fyco.wau.nl.  
Location/Qualifiers

#### FEATURES

##### source

1..438  
/organism="Phytophthora infestans"  
/mol\_type="mRNA"  
/strain="DDR7602, A1 mating type"  
/db\_xref="taxon:4787"  
/dev\_stage="4-week old vegetative, non-sporulating  
mycelium in synthetic medium"  
/lab\_host="E. coli, strain DHS-alpha"  
/clone\_lib="Plinfestammy"  
/note="Vector: pSPOR1; Site 1: SalI; Site 2: NotI; Total  
RNA was isolated from mycelium of P. infestans DDR7602  
cultured for 4 weeks in synthetic medium. EST clones were  
named by their position in the microtiter plate, preceded  
by the prefix MY (for mycelial) and the successive number  
of the microtiter plate (e.g. MY-06-A-04)."

##### ORIGIN

Query Match 94.1%; Score 16; DB 2; Length 438;  
Best Local Similarity 94.1%; Pred. NO. 3.1e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGGAC 17  
|||||  
Db 233 TCCGCCGGGTCCGGGAC 249

#### RESULT 3

##### LOCUS

BG857100 720 bp mRNA linear EST 29-MAY-2001  
DEFINITION 1024049H09.x1 C. reinhardtii CC-1690, normalized, lambda Zap II  
Chlamydomonas reinhardtii cDNA, mRNA sequence.

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

Chlamydomonas reinhardtii  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonadae.

##### REFERENCE

##### AUTHORS

##### TITLE

##### ANALYSES

##### UNICELLULAR

##### SYSTEM

##### FOR

##### ANALYZING

##### GENE

##### FUNCTION

##### AND

##### REGULATION

##### IN

##### VASCULAR

##### PLANTS;

##### PROJECT

##### PHASE

##### 2

##### UNPUBLISHED

##### (2000)

##### CONTACT:

##### CHARLES

##### HAUSER

##### DCMB

##### BOX

##### 91000

##### DURHAM,

##### NC

##### 27708-1000

##### TEL:

##### 919

##### 613

##### 8159

##### FAX:

##### 919

##### 613

##### 8177

##### EMAIL:

##### chauser@duke.edu.

#### FEATURES

##### source

1..720  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:1055"  
/clone\_lib="C. reinhardtii CC-1690, normalized, lambda Zap  
II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI; This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP (acetate-containing) medium in the  
light, TAP medium in the dark, HS (minimal) medium in  
ambient levels of CO2 and HS medium bubbled with 5% CO2.  
PolyA mRNA was purified from each sample, pooled and cDNA

synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

## ORIGIN

Query Match 94.1%; Score 16; DB 4; Length 720;  
Best Local Similarity 100.0%; Pred. No. 2.9e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCGCCGGGTCCGGGAC 17  
|||||  
Db 479 CCGCCGGGTCCGGGAC 494

RESULT 4  
LOCUS BU838754.C 932 bp mRNA linear EST 16-OCT-2002  
DEFINITION AGENCOURT 8210078 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6258013  
5' mRNA sequence.  
ACCESSION BU838754  
VERSION BU838754.1 GI:24023149  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 932)  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: DCTD/DRP  
CNA Library Preparation: Rubin Laboratory  
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM2413 row: d column: 14  
High quality sequence stop: 543.  
Location/Qualifiers

## FEATURES

## SOURCE

1. 932  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6258013"  
/tissue\_type="melanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_112"  
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGGAGG (g). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 94.1%; Score 16; DB 5; Length 932;  
Best Local Similarity 100.0%; Pred. No. 2.8e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCGCCGGGTCCGGGAC 17  
|||||  
Db 930 CCGCCGGGTCCGGGAC 915

RESULT 5

A1316193 193 bp mRNA linear EST 17-DEC-1998  
LOCUS A1316193  
DEFINITION u26b02.v1 Sugano mouse kidney mR1a Mus musculus cDNA clone  
IMAGE:1921035 5' similar to gb:X56295\_rnal PLASMA GLUTATHIONE  
PEROXIDASE PRECURSOR (HUMAN) ; gb:U13705 Mus musculus domesticus  
C57BL/6J plasma glutathione (mouse) ; mRNA sequence.

## ACCESSION

A1316193  
A1316193.1 GI:4031460  
EST  
Mus musculus (house mouse)

## SOURCE

Mus musculus

## REFERENCE

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:97327  
Seg primer: custom primer used  
High quality sequence stop: 192.  
Location/Qualifiers

## FEATURES

## SOURCE

1. 193  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1921035"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="Sugano mouse kidney mR1a"  
/note="Organ: kidney; Vector: pMB18S-FL3; Site 1: DraIII (CACTGTG); Site 2: DraIII (CACTGTG); 1st strand cDNA was primed with an oligo(dT) primer (ATGTGGCTTTTCTTTTCTTTT) ; double-stranded cDNA was ligated to a DraIII adaptor (TGTGGCTTACTGG) ; digested and cloned into distinct DraIII sites of the pMB18S-FL3 vector (5' site CACTGTG, 3' site CACTGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGCTCTTAAACTGG and 3' end primer GCAGCTGACGTGACGACA."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 1; Length 193;  
Best Local Similarity 94.1%; Pred. No. 6.7e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCCGGGAC 17  
|||||  
Db 96 TCCGCCGGGTCCGGGAC 112

## RESULT 6

LOCUS BY000112 217 bp mRNA linear EST 06-DEC-2002  
DEFINITION BY000112 RIKEN full-length enriched, adult male kidney Mus musculus  
cDNA clone 0610007p21 5', mRNA sequence.

FEATURES

ACCESSION	EF0000112
VERSION	EF0000112.1
KEYWORDS	GI:26060361
SOURCE	EST
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 217)
	Okazaki, Y., Furuno, M., Kankawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Oseto, N., Saito, R., Suzuki, H., Yamana, A., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schondache, C., Gotohori, T., Baldarilli, R., Hill, D. P., Butt, C., Hume, D. A., Quakebush, J., Schiml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Bratic, V., Chotia, C., Corbani, L. E., Cousins, S., Dalia, E., Dragan, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gasterland, T., Gariboldi, M., Giesi, C., Godzik, A., Gough, J., Grimmond, S., Gustintchik, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedziestki, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pereira, G., Perole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sendlin, A., Schneider, C., Semp, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wymshew-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Aizawa, T., Fukuda, S., Hara, A., Heshizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	123534681
PUBMED	12466851
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa, K., Akimura, T., Aizawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numata, K., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Watanabe, M., and Hayashizaki, Y. Direct Submission Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome, 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (http://genome.gsc.riken.go.jp) for further details. Location/Qualifiers

ORIGIN	Query Match	Best Local Similarity	Matches
1	90.6%;	94.1%;	16;
16;	Score 15.4;	Pred. No. 6.6e+03;	Conservative 0;
0;	DB 5;	Mismatches 1;	Indels 0;
0;	Length 217;	Gaps 0;	
1			
161	TCGCGCGGCGCGGAGAC 17	TCGCGCGGCGCGGAGAC 177	
DB			
RESULT 7			
LOCUS	BB561944	218 bp	RNA
DEFINITION	BB561944 RIKEN full-length enriched, adult male kidney Mus musculus		
ACCESSION	CNNA clone 0610005A12 5', mRNA sequence.		
VERSION	BB561944		
KEYWORDS	BB561944.1 GI:11452836		
SOURCE	EST.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Mus musculus		
AUTHORS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 218)		
	Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P., Haneagaki, T., Hayatsu, N., Hirooka, T., Hirozane, T., Hodojima, Y., Imotani, K., Ishii, Y., Itoh, M., Iwata, M., Kawai, J., Kojima, Y., Komori, K., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka, T., Toya, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.		
	RIKEN Mouse ESTs (Aizawa, K. et al. 2000)		
	Unpublished (2000)		
	Contact: Yoshihide Hayashizaki		
	Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute		
	The Institute of Physical and Chemical Research (RIKEN)		
	1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan		
	Tel.: 81-45-503-9222		
	Fax: 81-45-503-9216		
	Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/		
	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagao, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
	Thermotabilization and thermostabilization of full length		
	trachealose and its application for the synthesis of full length		
	cDNA. P roc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)		



Itch,M., Kitanai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,  
Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

#### FEATURES

##### source

Location/Qualifiers  
1..218  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="0610005A12"  
/sex="male"  
/tissue\_type="kidney"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/clone.lib="RIKEN full-length enriched, adult male kidney"  
/note="Site\_1: XhoI; Site\_2: SctI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5',  
GAGAGAGAGAGATCGAGCTCAATTAATTAATTAACCCCCCCCC 3'].  
cDNA was cleaved with XhoI and SctI. "

##### ORIGIN

Query Match 90.6%; Score 15.4; DB 2; Length 218;  
Best Local Similarity 94.1%; Pred. No. 6.6e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCGCCGGCTCGGGAC 17  
|||||||  
Db 97 TCCGCCGGCGCGGAC 113

RESULT 8 218 bp mRNA linear EST 06-DEC-2002  
BY000049  
LOCUS BY000049 RIKEN full-length enriched, adult male kidney Mus musculus  
DEFINITION cDNA clone 0610007122 5', mRNA sequence.  
ACCESSION BY000049  
VERSION BY000049.1 GI:26060298  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 218)  
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,  
Nikaido,I., Otsu,N., Saito,R., Suzuki,H., Yamanaka,I.,  
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,  
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bulc,C.,  
Hume,D.A., Quackenbush,J., Schiml,L.W., Kanapin,A., Matwuda,H.,  
Batilov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,  
Chochia,C., Corbani,L.E., Cousins,S., Dalla,E., Drisgani,T.A.,  
Fletcher,C.F., Forrest,A., Fraser,K.S., Gaasterland,T.,  
Gariboldi,M., Giasi,C., Godzik,A., Gough,J., Grimmond,S.,  
Gustlich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,  
Kawaji,H., Kawasawa,Y., Kedzierski,R.W., King,B.L., Konagaya,A.,  
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,  
Maltaise,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,  
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,

#### FEATURES

##### source

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Petrovsky,N., Pilaj,R., Pontius,J.U., Qi,D., Ramchandran,S.,  
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,  
Sundell,A., Schneider,C., Sempke,C.A., Setou,M., Shinada,K.,  
Gulana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,  
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,  
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,L.,  
Yang,L., Yuan,Z., Zavalan,M., Zhu,Y., Zimmer,A., Carninci,P.,  
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,  
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,  
Arikawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,  
Itoch,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,  
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
Rogers,J., Birney,E. and Hayashizaki,Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851  
22354683  
Contact: Yoshitake Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL:<http://genome.gsc.riken.jp/>  
Aizawa,K., Akimura,T., Arikawa,T., Carninci,P., Fukuda,S.,  
Hirozane,T., Imotani,K., Ishii,Y., Itoch,M., Kawai,J., Konno,H.,  
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
Shiraki,T., Tagami,M., Waki,K., Watanabe,A., Muramatsu,M. and  
Hayashizaki,Y. Direct Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

#### FEATURES

##### source

Location/Qualifiers  
1..218  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="0610007122"  
/sex="male"  
/tissue\_type="kidney"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/clone.lib="RIKEN full-length enriched, adult male kidney"  
/note="Site\_1: XhoI; Site\_2: SctI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5',

ORIGIN

GAGAGAGAGATCCAGAGCTCATTAATTAATTAACCCCCCCC 3'.  
cDNA was cleaved with XhoI and SstI. "

Query Match 90.6%; Score 15.4; DB 5; Length 218;  
Best Local Similarity 94.1%; Pred. No. 6.6e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TCCGCGGGGTCCGGGAC 17  
Db 98 TCCGCGGGGCGCGGAC 114

RESULT 9  
AY18508  
LOCUS  
DEFINITION Mus musculus GPX3 gene, VIRTUAL TRANSCRIPT, partial sequence.  
ACCESSION AY18508  
VERSION AY18508.1 GI:39774468  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

1 (bases 1 to 219)  
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tenenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J.,  
Adams,M.D. and Cargill,M.  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
Interfing nonneutral evolution from human-chimp-mouse orthologous  
gene trios

JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

2 (bases 1 to 219)  
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tenenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J.,  
Adams,M.D. and Cargill,M.  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
Interfing nonneutral evolution from human-chimp-mouse orthologous  
gene trios

COMMENT  
FEATURES  
source

1..219  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1..>219  
/gene="GPX3"  
/locus\_tag="Hm6580"

ORIGIN

Query Match 90.6%; Score 15.4; DB 9; Length 219;  
Best Local Similarity 94.1%; Pred. No. 6.6e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TCCGCGGGGTCCGGGAC 17  
Db 57 TCCGCGGGGCGCGGAC 73

RESULT 10  
BY006303  
LOCUS  
DEFINITION BY006303 RIKEN full-length enriched, adult male kidney Mus musculus  
cDNA clone F520009F08 5', mRNA sequence.  
ACCESSION BY006303  
VERSION BY006303.1 GI:26066552  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS

1 (bases 1 to 222)  
Okazaki,Y., Furuno,M., Kaekawa,T., Adachi,J., Bono,H., Kondo,S.,  
Nikaido,I., Otsu,N., Saito,R., Suzuki,H., Yamamoto,T.,  
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,  
Schonbach,C., Gojobori,T., Baldairelli,R., Hill,D.P., Bull,C.,  
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Macaula,H.,  
Batalov,S., Beisel,K.W., Blake,J.A., Bird,D., Brusic,V.,  
Chochua,C., Corbani,L.E., Cousine,S., Dalia,E., Dargatzis,T.A.,  
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,  
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,  
Guellinich,S., Hirokawa,N., Jackson,I.J., Jarvis,B.D., Kanai,A.,  
Kawai,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Kongsaya,A.,  
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,  
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,  
Nunata,K., Okido,T., Pavan,W.J., Petosa,G., Pezole,G.,  
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramchandran,S.,  
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,  
Sandelin,A., Schneider,C., Sempile,C.A., Setou,M., Shimada,K.,  
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,  
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Warande,Y.,  
Wells,C., Wilming,L.G., Wyszewski,B., Yanagisawa,M., Yang,I.,  
Yang,L., Yuan,Z., Zavoian,M., Zhu,Y., Zimmer,A., Carninci,P.,  
Hayatsu,N., Hirozane-Kishikawa,T., Kono,H., Nakamura,M.,  
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,  
Arakawa,T., Fukuda,S., Hara,A., Hashizume,M., Imokani,K., Ishii,Y.,  
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,  
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
Rogers,J., Birney,E. and Hayashizaki,Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

JOURNAL  
PUBMED  
REFERENCES  
AUTHORS

1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 220-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sec.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
Hirozane,T., Imokani,K., Ishii,Y., Itoh,M., Kawai,J., Kono,H.,  
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numasaki,R.,  
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
Shiraki,T., Tagami,M., Waki,K., Watanabe,A., Muramatsu,M. and  
Hayashizaki,Y. Direct Submision  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
Location/Qualifiers  
1..222  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="F520009F08"

FEATURES  
source

/sex="male"  
 /tissue.type="kidney"  
 /dev\_stage="adult"  
 /lab\_host="SOLR"  
 /clone\_lib="RIKEN full-length enriched, adult male kidney"  
 /note="Site 1: XhoI; Site 2: SctI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAGATCCAGAGCTCAATTAATTATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SctI."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 5; Length 222;  
 Best Local Similarity 94.1%; Pred. No. 6.6e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TCCGCGGGGTGCGGAC 17  
 |||||

Db 97 TCCGCGGGGTGCGGAC 113

## RESULT 11

BY006302 228 bp mRNA linear EST 06-DEC-2002  
 LOCUS BY006302  
 DEFINITION cDNA clone F520009F04 5', mRNA sequence.

ACCESSION BY006302.1 GI:2606551

VERSION EST.

KEYWORDS Mus musculus (house mouse)

## SOURCE

## ORGANISM

Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 228)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamataka, I.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,  
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,  
 Baralov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,  
 Chochia, C., Corbani, L. E., Cousine, S., Dalla, E., Dragani, T. A.,  
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,  
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
 Guitlich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,  
 Kawasaki, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,  
 Kurouchin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,  
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
 Numata, K., Okido, T., Pavan, W. J., Perce, G., Pesole, G.,  
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,  
 Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,  
 Saitana, R., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,  
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,  
 Vervari, R., Wagner, L., Wahlested, C., Wang, Y., Watanabe, Y.,  
 Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
 Yang, L., Yuan, Z., Zavoian, M., Zhu, Y., Zimmer, A., Carninci, P.,  
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imocani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,  
 Rogers, J., Birney, E. and Hayashizaki, Y.

Analyses of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

## MEDLINE

22354683  
 PUBMED 12466851

## COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Saitama-shi, Tsukuba-shi, Ibaraki-shi, Japan  
 Tel.: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
 Hirozane, T., Imocani, K., Ishii, Y., Itoh, M., Kawai, J., Kondo, H.,  
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numata, K.,  
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
 Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and  
 Hayashizaki, Y. Direct Substitution  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.

## FEATURES

## source

Location/Qualifiers  
 1..228

/organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="F520009F04"  
 /sex="male"  
 /tissue.type="kidney"  
 /dev\_stage="adult"  
 /lab\_host="SOLR"  
 /clone\_lib="RIKEN full-length enriched, adult male kidney"  
 /note="Site 1: XhoI; Site 2: SctI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAGATCCAGAGCTCAATTAATTATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SctI."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 5; Length 228;  
 Best Local Similarity 94.1%; Pred. No. 6.6e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## Oy

1 TCCGCGGGGTGCGGAC 17  
 |||||

## Db

99 TCCGCGGGGTGCGGAC 115

## RESULT 12

BB561733

BB561733

234 bp

mRNA

linear

EST 29-NOV-2000

DEFINITION BB561733 RIKEN full-length enriched, adult male kidney Mus musculus  
cDNA clone 0600001f01 5', mRNA sequence.

ACCESSION BB561733  
VERSION BB561733.1 GI:11452625  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 234)

REFERENCE Atawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,  
AUTHORS Carninci,P., Hanagaki,T., Hayatsu,N., Hirooka,T., Hirozane,T.,  
Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,  
Kojima,Y., Kono,H., Kusabe,M., Matsuyama,T., Miyazaki,A.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,  
Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,  
Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y.,  
Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,  
Watabiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S.,  
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermostabilization and thermocactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Tomaru,Y., Carninci,P., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
source location/Qualifiers  
1..234  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="0600001f01"  
/sex="male"  
/tissue\_type="kidney"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/clone\_1ib="RIKEN full-length enriched, adult male kidney"  
/note="Site 1: XhoI; Site 2: SctI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGCCGCGACGAGTGTGTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGCATCCAGAGCTCAATTAATTAATTAACCCCCCCCC 3'].  
cDNA was cleaved with XhoI and SctI. "

ORIGIN  
Query Match 90.6%; Score 15.4; DB 2; Length 234;

Best Local Similarity 94.1%; Pred. No. 6.5e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OR 1 TCCGCCGGGTGGGGAC 17  
DB 98 TCCGCCGGGTGGGGAC 114

RESULT 13  
LOCUS BB561740 236 bp mRNA linear EST 29-NOV-2000  
DEFINITION BB561740 RIKEN full-length enriched, adult male kidney Mus musculus  
cDNA clone 0600001g01 5', mRNA sequence.

ACCESSION BB561740  
VERSION BB561740.1 GI:11452632  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 236)

REFERENCE Atawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,  
AUTHORS Carninci,P., Hanagaki,T., Hayatsu,N., Hirooka,T., Hirozane,T.,  
Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,  
Kojima,Y., Kono,H., Kusabe,M., Matsuyama,T., Miyazaki,A.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,  
Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,  
Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y.,  
Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,  
Watabiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S.,  
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermostabilization and thermocactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Tomaru,Y., Carninci,P., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
source location/Qualifiers  
1..236  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="0600001g01"  
/sex="male"  
/tissue\_type="kidney"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/clone\_1ib="RIKEN full-length enriched, adult male kidney"  
/note="Site 1: XhoI; Site 2: SctI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'  
GAGAGAGAGCGCCGACACTCGATTGTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGAGATCAAGAGCTCAATTAATTAAATTAACCCCCCCCC 3'].  
cDNA was cleaved with XhoI and SstI."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 2; Length 236;  
Best Local Similarity 94.1%; Pred. No. 6.5e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCGGGAC 17  
|||||  
Db 99 TCCGCCGGGTCGGGAC 115

## RESULT 14

AU096437 243 bp mRNA linear EST 03-APR-2002  
LOCUS AU096437 Rice green shoot Oryza sativa (japonica cultivar-group)  
DEFINITION CDNA clone SJ3687, mRNA sequence.  
ACCESSION AU096437  
VERSION AU096437.1 GI:8859119  
KEYWORDS EST.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzoae; Oryza.  
REFERENCE 1 (bases 1 to 243)  
Sasaki, T. and Yamamoto, K.  
Rice cDNA from green shoot (2000)  
Unpublished (2000)  
TITLE Rice cDNA from green shoot (2000)  
JOURNAL Unpublished (2000)  
CONTACT: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
305-8602, Japan  
Tel: 81-298-38-7441  
Fax: 81-298-38-7468  
Email: tsasaki@abr.affrc.go.jp, URL: http://rsgp.dna.affrc.go.jp/  
PROJECT = "RGP".

## FEATURES

source location/Qualifiers  
1..243  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="SJ3687"  
/clone\_1ib="Rice green shoot"  
/note="Green shoot (8 days old)"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 1; Length 243;  
Best Local Similarity 94.1%; Pred. No. 6.5e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCGGGAC 17  
|||||  
Db 115 TCCGCCGGGTCGGGAC 131

## RESULT 15

AA277884 244 bp mRNA linear EST 01-APR-1997  
LOCUS AA277884 vcl0c05.r1 Bartsch MRLB1 Mus musculus cDNA clone IMAGE:766088 5'  
DEFINITION similar to gb:U3705 Mus musculus domesticus C57BL/6J plasma  
glutathione (MOUSE);, mRNA sequence.  
ACCESSION AA277884  
VERSION AA277884.1 GI:1918051  
KEYWORDS EST.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

## REFERENCE

1 (bases 1 to 244)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Giesel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMNI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:467008  
Seq primer: -28m3 rev2 ET from Amersham  
High quality sequence stop: 209.

## TITLE

The WashU-HMNI Mouse EST Project  
Unpublished (1996)

## JOURNAL

Unpublished (1996)

## COMMENT

Unpublished (1996)

## FEATURES

source location/Qualifiers  
1..244  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
/clone="IMAGE:766088"  
/sex="mixed"  
/tissue\_type="Kidney"  
/dev\_stage="6 weeks"  
/lab\_host="DH10B"  
/clone\_1ib="Bartsch MRLB1"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCAGATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT  
3'] double-stranded cDNA was ligated to Eco RI adaptors  
[CATGATTCGTAAC], digested with Not I and cloned into the  
Not I and Eco RI sites of the modified pT73 vector.  
Library constructed by Bob Bartsch."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 1; Length 244;  
Best Local Similarity 94.1%; Pred. No. 6.5e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCGGGAC 17  
|||||  
Db 31 TCCGCCGGGTCGGGAC 47

Search completed: November 7, 2005, 09:18:34  
Job time : 1704.72 secs

***This page under review***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:11:16 ; Search time 785.957 Seconds  
(without alignments) 1233.024 Million cell updates/sec

Title: US-10-777-131a-3

Perfect score: 20

Sequence: 1 tcgaagtggtggaatcgtg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 2422767955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sbs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	20	100.0	20	AX770105	AX770105 Sequence
2	20	100.0	1144	AX770122	AX770122 Sequence
3	20	100.0	35336	MSGY367	AD000008 Mycobacte
4	20	100.0	103837	AE000516_43	Continuation (44 O
5	20	100.0	244800	BX842584	BX842584 Mycobacte
6	20	100.0	278492	BX248347	BX248347 Mycobacte
7	18	90.0	349895	BX248359	BX248359 Corynebac
8	17.4	87.0	143966	AC135920	AC135920 Oryza sat
9	17.4	87.0	204254	AC122848	AC122848 Mus muscu
10	17	85.0	176823	AC146138	AC146138 Pan trogl
11	17	85.0	178455	AC146015	AC146015 Pan trogl
12	16.8	84.0	254554	CEY102A5D	AL031628 Caenorhab
13	16.8	84.0	108424	AL669827	AL669827 Mouse DNA
14	16.8	84.0	147180	AC136469	AC136469 Homo sapi
15	16.8	84.0	156942	AC136470	AC136470 Homo sapi
16	16.8	84.0	166618	AC010323	AC010323 Homo sapi
17	16.8	84.0	168982	AC125324	AC125324 Mus muscu
18	16.8	84.0	189562	AC148560	AC148560 Oryctolag
19	16.8	84.0	196855	AC022297	AC022297 Mus muscu

C	20	16.8	84.0	202232	2	AC124759	AC124759 Mus muscu
C	21	16.8	84.0	204037	10	AL596207	AL596207 Mouse DNA
C	22	16.8	84.0	211892	10	AL732403	AL732403 Mouse DNA
C	23	16.8	84.0	228645	10	AL450397	AL450397 Mouse DNA
C	24	16.8	84.0	231611	2	AC104870	AC104870 Mus muscu
C	25	16.8	84.0	235041	2	AC130100	AC130100 Rattus no
C	26	16.8	84.0	249984	2	AC139594	AC139594 Rattus no
C	27	16.4	82.0	1890	1	AF352014	AF352014 Vibrio gp
C	28	16.4	82.0	11803	1	AB015554	AB015554 Shewanell
C	29	16.4	82.0	33896	1	AY523728S	AY523728 Azospirill
C	30	16.4	82.0	44226	9	AY130859	AY130859 Homo sapi
C	31	16.4	82.0	44300	2	AC006109	AC006109 Homo sapi
C	32	16.4	82.0	48703	2	AC100448	AC100448 Homo sapi
C	33	16.4	82.0	69399	8	AP006350	AP006350 Lotus cor
C	34	16.4	82.0	87645	1	AP006659	AP006659 Lotus cor
C	35	16.4	82.0	110000	1	AP006841_12	AP006841 Continuation (13 O
C	36	16.4	82.0	111311	9	HS27919	AL033517 Human DNA
C	37	16.4	82.0	125980	9	AC093223	AC093223 Homo sapi
C	38	16.4	82.0	150263	9	AC145132	AC145132 Homo sapi
C	39	16.4	82.0	152464	2	AC027421	AC027421 Homo sapi
C	40	16.4	82.0	157243	9	AL353622	AL353622 Human DNA
C	41	16.4	82.0	157986	2	AC025860	AC025860 Homo sapi
C	42	16.4	82.0	164597	3	AC011907	AC011907 Drosophila
C	43	16.4	82.0	167178	2	AC025672	AC025672 Homo sapi
C	44	16.4	82.0	170575	2	AC019758	AC019758 Drosophila
C	45	16.4	82.0	170977	9	AP000756	AP000756 Homo sapi

#### ALIGNMENTS

RESULT 1	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
DEFINITION	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
ACCESSION	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
VERSION	AX770105.1	GI:32437683	20 bp	DNA	linear	PAT 02-JUL-2003
KEYWORDS						
SOURCE						
ORGANISM						
synthetic construct						
other sequences; artificial sequences.						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
INSTITUT						
PATENT						
DATE						
FEATURES						
source						
1..20						
/organism="synthetic construct"						
/mol_type="unassigned DNA"						
/db_xref="taxon:32630"						
/note="Primer"						
ORIGIN						
Query Match						
Beet Local Similarity						
Matches						
20; Conservative						
0; Mismatches						
0; Indels						
0; Gaps						
0;						
QY						
1						
TCGAAGTGGGCAATCGTG						
20						
DB						
1						
TCGAAGTGGGCAATCGTG						
20						
RESULT 2						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
Mycobacterium tuberculosis						
Mycobacterium tuberculosis						

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE  
1 Glouquel, B.  
Compositions and methods for detecting multidrug resistant strains  
of *M. tuberculosis* having mutations in genes of the *mtr* family  
Patent: WO 03016562-A 20 27-FEB-2003;  
INSTITUT PASTEUR (FR)

JOURNAL  
location/Qualifiers

FEATURES  
Source  
1. 1144  
/organism="Mycobacterium tuberculosis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1773"

ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCGAAGTGGGCAATCGTG 20  
7 TCGAAGTGGGCAATCGTG 26

Db

RESULT 3  
MSGY367 35336 bp DNA linear BCT 03-DEC-1996  
LOCUS  
DEFINITION Mycobacterium tuberculosis sequence from clone y367.  
ACCESSION  
AD000008  
VERSION  
AD000008.1 GI:1702971

SOURCE  
Mycobacterium tuberculosis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE  
1 (bases 1 to 35336)  
Du, L.  
Direct Submission  
Submitted (11-OCT-1996) L. Du, Genome Therapeutics Corporation, 100  
Beaver Street, Waltham, MA, USA, 02154 du@ctic.com

COMMENT  
GSDB:S:1004717.

FEATURES  
Source  
1. 35336  
/organism="Mycobacterium tuberculosis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1773"  
/clone="y367"

ORIGIN  
Query Match 100.0%; Score 20; DB 1; Length 35336;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCGAAGTGGGCAATCGTG 20  
14705 TCGAAGTGGGCAATCGTG 14686

Db

RESULT 4  
AE000516\_43  
WPCOMMENT  
Sequence split into 44 fragments LOCUS AE000516 Accession AE000516

Fragment Name Begin End  
AE000516\_00 1 110000  
AE000516\_01 100001 210000  
AE000516\_02 200001 310000  
AE000516\_03 300001 410000  
AE000516\_04 400001 510000  
AE000516\_05 500001 610000  
AE000516\_06 600001 710000  
AE000516\_07 700001 810000

AE000516\_08 800001 910000  
AE000516\_09 900001 1010000  
AE000516\_10 1000001 1110000  
AE000516\_11 1100001 1210000  
AE000516\_12 1200001 1310000  
AE000516\_13 1300001 1410000  
AE000516\_14 1400001 1510000  
AE000516\_15 1500001 1610000  
AE000516\_16 1600001 1710000  
AE000516\_17 1700001 1810000  
AE000516\_18 1800001 1910000  
AE000516\_19 1900001 2010000  
AE000516\_20 2000001 2110000  
AE000516\_21 2100001 2210000  
AE000516\_22 2200001 2310000  
AE000516\_23 2300001 2410000  
AE000516\_24 2400001 2510000  
AE000516\_25 2500001 2610000  
AE000516\_26 2600001 2710000  
AE000516\_27 2700001 2810000  
AE000516\_28 2800001 2910000  
AE000516\_29 2900001 3010000  
AE000516\_30 3000001 3110000  
AE000516\_31 3100001 3210000  
AE000516\_32 3200001 3310000  
AE000516\_33 3300001 3410000  
AE000516\_34 3400001 3510000  
AE000516\_35 3500001 3610000  
AE000516\_36 3600001 3710000  
AE000516\_37 3700001 3810000  
AE000516\_38 3800001 3910000  
AE000516\_39 3900001 4010000  
AE000516\_40 4000001 4110000  
AE000516\_41 4100001 4210000  
AE000516\_42 4200001 4310000  
AE000516\_43 4300001 4400000  
Continuation (44 of 44) of AE000516 from base 4300001 (AE000516 Mycobacterium tuberculosis)

Query Match 100.0%; Score 20; DB 1; Length 103837;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCGAAGTGGGCAATCGTG 20  
85562 TCGAAGTGGGCAATCGTG 85581

Db

RESULT 5  
BX842584  
LOCUS  
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 13/13.  
ACCESSION  
BX842584 AL021456 AL022076 AL022120 AL022121 AL123456 Z80343 Z83864  
Z94121 Z97188  
VERSION  
BX842584.1 GI:38490370  
KEYWORDS  
complete genome.  
SOURCE  
Mycobacterium tuberculosis H37Rv  
ORGANISM  
Mycobacterium tuberculosis H37Rv  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE  
1  
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,  
Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
Connor, R., Davies, R., Devlin, K., Fellwell, T., Gentles, S.,  
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
Squares, S., Squares, R., Sultson, J.E., Taylor, K., Whitehead, S. and  
Barrell, B.G.  
Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence  
Nature 393 (6685), 537-544 (1998)



MEDLINE 98295987  
PUBMED 9634230  
REFERENCE 2  
AUTHORS Camus,J.C., Pryor,M.J., Medigue,C. and Cole,S.T.  
TITLE Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv  
JOURNAL Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)  
MEDLINE 22255591  
PUBMED 12368430  
REFERENCE 3 (bases 1 to 244800)  
AUTHORS Parkhill,J.  
TITLE Direct Submission  
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterieme, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
COMMENT On or before Nov 21, 2003 this sequence version replaced  
gi:3261511, gi:32615026, gi:3261558, gi:3261559, gi:3261648,  
gi:3261687, gi:3261736, gi:3261805.  
NOTES:  
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)).  
Location/Qualifiers  
1. 244800  
/organism="Mycobacterium tuberculosis H37Rv"  
/mol\_type="genomic DNA"  
/strain="H37Rv"  
/db\_xref="taxon:83332"  
/complement(89..1396)  
/locus\_tag="RV3722C"  
/complement(89..1396)  
/locus\_tag="RV3722C"  
/function="UNKNOWN"  
/note="RV3722C, (MTV025..070C), len: 435 aa. Conserved hypothetical protein, equivalent to O69513|MCB2407.14 (alias O9CB73|ML2336, 463 aa) HYPOTHETICAL 46.8 KDA PROTEIN from Mycobacterium leprae (426 aa), FASTA scores: opt: 2505, E(): 8.3e-154, (67.25% identity in 424 aa overlap). Also highly similar to Q9RU17|DR1579 CONSERVED HYPOTHETICAL PROTEIN from Deinococcus radiodurans (452 aa), FASTA scores: opt: 1162, E(): 3.1e-67, (44.8% identity in 422 aa overlap); and partially similar to Q91371|PA1654 PROBABLE AMINOTRANSFERASE from Pseudomonas aeruginosa (388 aa) FASTA scores: opt: 162, E(): 0.0078, (25.85% identity in 348 aa overlap) and other aminotransferases. Tbpase score is 0.900. N-terminus extended since first submission (previously 408 aa)."  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CA18044.2"  
/db\_xref="GI:38490371"  
/translation="MSPSLSROELALAHHQGVYALQGMKLADLRGKSAEC  
DISNQLSPEDVDYRDEPTDTRNGGCHGPGALPAIFAEILGIVAPMLIAGNSSE  
LMDHIVASMLYGVDSVPWVIOEDQIKFLCPVGYDHFPAITETGIEMLPIPMQ  
DMDVDLIEELVADPAIKGMVTVFPGNPGSVTSWETVRLVQMTAAPFLFMD  
NAYAVHTLTLDFPRQVDVIGLAKAGNNRPFVFASTSKITPAGGVSPFGSLNIA  
WYLOYAGKSGISPDKNOLRHLRFEGDADGVRLHMLRHOQLAPKFAVAVELDRLS  
ESKIASWTEPKGVFISIDVLPCTARTVALAKDGIATVETAGAFPPRKDDDNIR  
IAPSPSPDRLNANDGLATCALLAATETLNLQGLASSANPVR"  
1613..1698  
/gene="cRNA-Ser (GGA)"  
/gene="cRNA-Ser (GGA)"  
1613..1698  
/gene="cRNA-Ser (GGA)"  
/product="cRNA-Ser"  
/note="codon recognized: UCC"  
/anticodon=(pos:1647..1649,aa:Ser)  
1804..2568  
/locus\_tag="RV3723"  
1804..2568  
CDS

gene  
/locus\_tag="RV3723"  
/function="UNKNOWN"  
/note="RV3723, (MTV025..071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to O69512|ML2337|MCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). Tbpase score is 0.900."  
/codon\_start=1  
/transl\_table=11  
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"  
/protein\_id="CA18045.1"  
/db\_xref="GI:2960147"  
/translation="MGRVAVLMAHSAFSGAGVLYFVLPWPELMDDTGHSLTGL  
RIATGALGLALPVFTLLTRRELETQPOLASMRWSTMAHVALALVTAISE  
VMLSIDAGQWLFGYGAALAVGFPFYLSPFAELPPPKPLKPKPKORLRR  
KTKAKGDAEPEAEAEANTLELAQDEAVAEAPESIESGEPESATREAPAETA  
TAEPREGLRRRRPTGKTSRRRRTRSGVQVAKVDE"  
2735..2977  
/gene="cut5a"  
/locus\_tag="RV3724A"  
2735..2977  
/gene="cut5a"  
/locus\_tag="RV3724A"  
/EC\_number="3.1.1.-"  
/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)."  
/note="RV3724A, (MTV025..072), len: 80 aa. Probable cut5a, truncated cutinase precursor (EC 3.1.1.-), similar to N-terminal end of others e.g. O9KH87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 202, E(): 1.5e-06, (56.45% identity in 62 aa overlap); Q9XB09|RVND2-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa), FASTA scores: opt: 200, E(): 1.5e-06, (61.4% identity in 57 aa overlap); and Q00298|CUT1\_BOTCI|CUT1A CUTINASE PRECURSOR from Botrytis cinerea (Botryotinia fuckeliana) (202 aa), FASTA scores: opt: 108, E(): 2.2, (40.4% identity in 52 aa overlap). Also highly similar to others from Mycobacterium tuberculosis e.g.  
O06318|CUT1\_MYCTU|RV3451|MT3557|MYCY13E12.04 PROBABLE CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 189, E(): 1.2e-05, (58.0% identity in 50 aa overlap);  
O50664|CUT1\_MYCTU|RV2301|MT2356|MYCY339.08C PROBABLE CUTINASE PRECURSOR (219 aa), FASTA scores: opt: 172, E(): 0.00015, (59.2% identity in 49 aa overlap);  
O06793|RV1758|MYCY28.24|Z95890 HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29, (57.2% identity in 166 aa overlap);  
O06319|RV3452|MYT13E12.05, and U00015..11 from Mycobacterium leprae. BELONGS TO THE CUTINASE FAMILY. Rest of cutinase ORF continues as RV3724B|CUT5B, frameshifting could occur near position 416968. Sequence has been checked but no errors found."  
/codon\_start=1  
/transl\_table=11  
/product="PROBABLE CUTINASE PRECURSOR [FIRST PART] CUT5A"  
/protein\_id="CAE55632.1"  
/db\_xref="GI:38490372"  
/translation="MDVIRMRRLAAVAGTAATTPGLLSAHVPMVASPCPDVEV  
PARCTGPRPGISGVGLFVDLARPPKRVYRGILRR"  
2874..3437  
/gene="cut5b"  
/locus\_tag="RV3724B"  
2874..3437  
/gene="cut5b"  
/locus\_tag="RV3724B"  
/EC\_number="3.1.1.-"  
/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)."  
/note="RV3724B, (MTV025..072), len: 187 aa. Probable cut5b, truncated cutinase (EC 3.1.1.-), similar to C-terminal end

of others e.g. Q9XB09|RV2D-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa) FASTA scores: opt: 33.5, E(): 3.4e-12 (53.25% identity in 92 aa overlap); Q9XB87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 25.1, E(): 2.5e-07, (44.05% identity in 168 aa overlap). Also similar to proteins from Mycobacterium tuberculosis e.g. Q06793|RV1758|MTCY28.24 HYPOTHEICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 64.1, E(): 2.5e-29, (57.25% identity in 166 aa overlap); Q06319|RV3452|MTCY1312.05 HYPOTHEICAL 23.1 KDA PROTEIN (225 aa), FASTA scores: opt: 38.5, E(): 7.5e-15, (46.65% identity in 165 aa overlap); Q06318|CUT3|MYCTU|RV3451|MT3557|MTCY1312.04 PROBABLE CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 30.7, E(): 1.9e-10, (40.7% identity in 167 aa overlap); Q10837|CUT1|MYCTU|RV1984C|MT2037|MTCY39.35 PROBABLE CUTINASE PRECURSOR (217 aa), FASTA scores: opt: 26.1, E(): 6.7e-08, (50.9% identity in 169 aa overlap); etc; and U00015.11 from Mycobacterium lepra. 5'-end of gene is RV3724A|CUT5A; frameshifting may occur near position 4169668. TBPase score is 0.918."

/codon\_start=1  
/transl\_table=1  
/product="PROBABLE CUTINASE [SECOND PART] CUT5B"  
/protein\_id="CAE55633.1"  
/db\_xref="GI:38490373"  
/translation="MAPGSHLVLAASEDCSTHCVSQVGAKSLGVAVNYPASNDPAS SDPKTYVDIGRIDAGSHISQWASGCPOTROYLSQGAAYAGVTSVAVPAPVQVA VPAPMAPVNAHVAAVTLFGAPSAQFQGYAPPAIAGPLQYPTKLQCADGDSICGD GNSPVAHGLVAVNMGVQGANFASRL"

Query Match 100.0%; Score 20; DB 1; Length 244800;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCGAAGGTGGCGAAATCGTG 20  
|||||  
Db 226523 TCGAAGGTGGCGAAATCGTG 226542

RESULT 6  
BX248347  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCES  
AUTHORS  
JOURNAL

BX248347 278492 bp DNA linear BCT 06-MAY-2004  
Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 14/14.  
BX248347 BX248333  
Complete genome.  
Mycobacterium bovis AF2122/97  
Mycobacterium bovis AF2122/97  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacteriales; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
1  
Garnier, T., Biglmeier, K., Camus, J.-C., Medina, N., Mansoor, H.,  
Pryor, M., Duboy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S.,  
Harris, B., Actin, R., Doggett, D., Mayes, R., Keating, L.,  
Wheeler, P. R., Parkhill, J., Barrrell, B. G., Cole, S. T., Gordon, S. V. and  
Hewinson, G.  
The complete genome sequence of Mycobacterium bovis  
Unpublished  
2 (bases 1 to 278492)  
Garnier, T.  
Direct Submission  
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire  
Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex  
15, France. e-mail: cgarnier@pasteur.fr Submitted on behalf of the  
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary  
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,  
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK. P4 Annotation, Genopole, Institut  
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28

FEATURES  
source  
gene  
CDS  
gene  
CDS  
gene  
CDS

ru du Docteur Roux, 75724 Paris Cedex 15, France  
Location/Qualifiers  
1..278492  
/organism="Mycobacterium bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:333413"  
complement(240..704)  
/locus\_tag="MB3713c"  
complement(240..704)  
/locus\_tag="MB3713c"  
/note="MB3713c", len: 154 aa. Equivalent to RV3686C.  
len: 154 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 154 aa overlap). Hypothetical protein,  
similar to other bacterial hypothetical proteins e.g.  
Q9X341|SC166.02c from Streptomyces coelicolor (154 aa),  
FASTA scores: opt: 42.5, E(): 3.4e-19, (46.1% identity in  
154 aa overlap); Q9WZP4|TM0690 from Thermotoga maritima  
(149 aa), FASTA scores: opt: 32.6, E(): 3.4e-13, (40.4%  
identity in 151 aa overlap); Q9PHU3|CU0573 from  
Campylobacter jejuni (147 aa), FASTA scores: opt: 29.0,  
E(): 5.1e-11, (36.4% identity in 151 aa overlap); etc.  
Also some similarity to upstream  
Q63654|RV3686C|MTV025.034c CONSERVED HYPOTHETICAL PROTEIN  
from Mycobacterium tuberculosis."

/codon\_start=1  
/transl\_table=1  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD95899.1"  
/db\_xref="GI:31620457"  
/db\_xref="UniProt/TREMBL:Q7TVX7"  
/translation="MAELKSQLRDLDTQAMKTDQDKLTATIRMLAAIQTGEVSGKQA  
RELSDDEVIKTLARESRKRGAAEIVYONGGEAAEHAARIDEVLPPLTEGEL  
ADVDVTAIAEVAEELGHRPSKMGVLWKATVIAAGADGARLSAAVKERL"  
704..2059  
/locus\_tag="MB3714"  
704..2059  
/locus\_tag="MB3714"  
/locus\_tag="MB3714", len: 451 aa. Equivalent to RV3689, len:  
451 aa, from Mycobacterium tuberculosis strain H37Rv,  
(99.8% identity in 451 aa overlap). Probable conserved  
transmembrane protein, with proline rich N-terminus,  
similar to Q9K1W6|SC33.17 PUTATIVE INTEGRAL MEMBRANE  
PROTEIN from Streptomyces coelicolor (462 aa) FASTA  
scores: opt: 73.0, E(): 2.7e-21, (38.1% identity in 412 aa  
overlap)."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"  
/protein\_id="CAD95900.1"  
/db\_xref="GI:31620458"  
/db\_xref="UniProt/TREMBL:Q7TVX6"  
/translation="MHKRYAPQRPKPPETVYIKCTDRRODGGHDEROLLRPVMLP  
PGYVEPPVAPGAYAPAGYPPYATPPYATPPYATPPYATPPYATPPYATPPYATPPY  
GVVPLQIISIALGPGMTAFDITVGEDELTGAVGWSASPASLIVSDAGLLS  
AAALVIAAGAFGSPITVGEAMVAVRGLALFGALAEAGVAVGLAVIISGVA  
GAANEAAALGFPPLLVGSLAVLVLLVLFALLVLEPLVLEALTRFALVRG  
PMVLTGIRLTVLVGVGNALAPFMIIVGIVVAVVASDGSVTRLCATGLSAIVGT  
IGQIVTAPFSAGVVLLVTDRIARAFAFDVLQGLGEPGAPGAPVESTNMLVLRP  
P"  
2086..2739  
/locus\_tag="MB3715"  
2086..2739  
/locus\_tag="MB3715"  
/note="MB3715", len: 217 aa. Equivalent to RV3679, len:  
217 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 217 aa overlap). Probable conserved  
membrane protein, similar to Q9K1W5|SC33.18 PUTATIVE  
INTEGRAL MEMBRANE PROTEIN from Streptomyces coelicolor  
(231 aa), FASTA scores: opt: 41.9, E(): 1.5e-19, (36.0%  
identity in 211 aa overlap). Equivalent to AAK48159 from  
Mycobacterium tuberculosis strain CDC1551 (233 aa) but

gene

CDS

5073. .6395  
/locus\_tag="Mb3718"  
5073. .6395  
/locus\_tag="Mb3718"  
/note="Mb3718"  
/note="Mb3718". len: 440 aa. Equivalent to R3693, len: 440 aa (alternative start at 41910), from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 440 aa overlap). Possible conserved membrane protein, similar to O9KWM2|SC633.21 PUTATIVE LIPOPROTEIN from Streptomyces coelicolor (436 aa), FASTA scores: opt: 875, E(): 3.3e-46, (56.25% identity in 448 aa overlap), Q9NAN0|CC0567 HYPOTHETICAL PROTEIN from Caulobacter crescentus (437 aa), FASTA scores: opt: 355, E(): 2.3e-14, (30.9% identity in 450 aa overlap), P73233|SLR2013 HYPOTHETICAL 48.5 KDA PROTEIN from Synecchocystis sp. strain PCC 6803 (435 aa), FASTA scores: opt: 340, E(): 1.9e-13, (29.7% identity in 438 aa overlap); etc. Equivalent to AAK48162 from Mycobacterium tuberculosis strain CDC151 (475 aa) but shorter 35 aa. Also similar to other hypothetical proteins from Mycobacterium tuberculosis; MT0014\_7; MT0007\_27; and

Query Match

Best Local Similarity 100.0%; Score 20; DB 1; Length 278492;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 260216 TCGAAGGTGGCAATCGTG 260235

1 TCGAAGGTGGCAATCGTG 20  
|||||  
|||||

RESULT 7  
LOCUS BX248359  
DEFINITION Corynebacterium diphtheriae gravis NCTC13129, complete genome;  
segment 6/8.  
ACCESSION BX248359  
VERSION BX248359.1  
KEYWORDS GI:38200531  
SOURCE complete genome.  
ORGANISM Corynebacterium diphtheriae  
Corynebacterium diphtheriae  
Corynebacterium diphtheriae  
Actinobacteria; Actinobacteriales; Corynebacteriaceae; Corynebacterium.  
1 (bases 1 to 349895)  
Cordenso-Tarraga,A.M., Eferatiou,A., Dover,L.G., Holden,M.T.G.,  
Pallen,M., Bentley,S.D., Beesa,G.S., Churcher,C., James,K.D., De  
Zoyza,A., Chillingworth,T., Cronin,A., Dowd,L., Felwell,T.,  
Hamlin,N., Holtroyd,S., Jagels,K., Moulé,S., Quail,M.A.,  
Rabinowitzsch,E., Rutherford,K., Thomson,N.R., Unwin,L.,  
Whitehead,S. and Barrall B.G. Parkhill,J.  
The complete genome sequence and analysis of Corynebacterium  
diphtheriae NCTC13129  
Nucleic Acids Res. 31 (22), 6516-6523 (2003)  
14602910  
2 (bases 1 to 349895)  
Cordenso-Tarraga,A.M.  
Direct Submission  
Submitted (03-OCT-2003) Cordenso-Tarraga A.M., submitted on behalf  
of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust  
Genome Campus, Hinxton, Cambridge CB10 1SA E-mail:  
amc@anger.ac.uk  
Location/Qualifiers  
1. .349895  
/organism="Corynebacterium diphtheriae"  
/mol\_type="genomic DNA"  
/strain="NCTC13129"  
/db\_xref="taxon:1171"  
/note="biotype gravis"  
91. .648  
/locus\_tag="DIP1698"  
91. .648

FEATURES  
source

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

gene

CDS

91. .648

```
/locus_tag="DIP1698"
/note="Similar to Corynebacterium glutamicum thiamine
biosynthesis protein X Thix SW:THX_CORCU (P42461) (33
aa) fasta scores: E(): 2.5e-12, 38.09% id in 126 aa"
/codon_start=1
/transl_table=11
/product="putative thiamine biosynthesis related protein"
/protein_id="CAE50227.1"
/db_xref="GI:38200532"
/translation="MTIRIRSLTTAAVLAALGAIACSPPHQVSDTVTRATSK
VAPSSASASAMAAPSDKNAETIPASATEMLTNGONTVELSGNPEVGYAAICNAT
TPGAPAPLCTGGADTSQAAMIKADGTGTALTNDGTATVNLTAATGQGVNCKTDK
CVIKVFGDHTGEGRFVTEVPVTFAS"
91..189
/locus_tag="DIP1698"
/note="Signal peptide predicted for DIP1698 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 0.728 between residues 33 and 34;
signal-peptide site"
337..360
/locus_tag="DIP1698"
/note="ScanRegExp hit to P500687, Aldehyde dehydrogenases
glutamic acid active site."
complement(653..700)
/note="Score 63: 21/21 (100%) matches, 0 gaps"
/gene="dnaG"
/locus_tag="DIP1699"
/complement(697..2595)
/gene="dnaG"
/locus_tag="DIP1699"
/BC_number="2.7.7.-"
/note="Similar to Mycobacterium tuberculosis DNA primase
DnaG or RV2343c or MT2408 or MTCY98.12c SW:PRIM_MYCTU
(P95239) (639 aa) fasta scores: E(): 1.4e-146, 59.11% id
in 631 aa, and to Escherichia coli DNA primase DnaG or
DnaP or ParB or B3066 or Z4419 or EC63949 SW:PRIM_ECOLI
(P02923) (581 aa) fasta scores: E(): 2e-38, 32.27% id in
505 aa"
/codon_start=1
/transl_table=11
/product="DNA primase"
/protein_id="CAE50228.1"
/db_xref="GI:38200533"
/translation="MAGKGRIPDSIDIAIRERTPIREIVGVOLKPAQDSLKGLSP
FKERTSPFHYRPNHGFHCGSTGKGVPEFLMEMELTPEAVCAEAKGQINQY
QGTTGARREEPGTRKRLIEANRYAHQFYRQOLEPEAAVARQFLDGFSDQHLYGF
BCGYAPBGMDDTKLFKMYNTPETELKXSKYLFGLDLAKSIAEQHVVVEGYTDV
IGFARLKLFDKDKLKYNTPELTKYKSLVPLGLDLAKSIAEQHVVVEGYTDV
AMHAAGITTAVAACGTAFGEHSLQILRRYMLDDVAFASLIYTFDGDGAQKAMRAP
EGDOKFGOSYVAAPDGMBCDLRLAKGDAVDVAVQRPMEFTIRTLISYPLD
SVEGRQLARRAIVASIRDPALKTETARILLAGVCGPDPEVLIQVRAERRPKTO
PRQRATRIDNSQPMATQSMIVMDREPRMLPQREALKIALQTPHAGFVPSLPD
DAFTHESYAAVRAKIVAAGCSDAIRGSTWMDAVADNYDVAGKSLVSELAVEHILID
EMGLAHTDTTLARMEQRIGNQIAQLKGAAGQMRMPSBDMQAVNSLFDVALRQTRR
ELKELIKPK"
complement(1546..1806)
/locus_tag="dnaG"
/locus_tag="DIP1699"
/note="HMPfam hit to PF01751, Toprim domain"
complement(1570..1806)
/gene="dnaG"
/locus_tag="DIP1699"
/note="HMMsmat hit to SM00493, No definition"
complement(2281..2577)
/gene="dnaG"
/locus_tag="DIP1699"
/note="HMMpfam hit to PF01807, CHC2 zinc finger"
complement(2311..2478)
/gene="dnaG"
/locus_tag="DIP1699"
/note="BlastProdom hit to PD002988, PD002988"
complement(2317..2481)
```

```
/gene="dnaG"
/locus_tag="DIP1699"
/note="HMMsmat hit to SM00400, zinc finger"
complement(2606..4483)
/gene="glnS"
/locus_tag="DIP1700"
/complement(2606..4483)
/gene="glnS"
/locus_tag="DIP1700"
/BC_number="2.6.1.16"
/note="Similar to Streptomyces coelicolor
glucosamine--fructose-6-phosphate aminotransferase
[isomerizing] GlnS or SC664.18 SW:GlnS_STRO (O08678) (614
aa) fasta scores: E(): 2.7e-154, 65.86% id in 624 aa, and
to Escherichia coli glucosamine--fructose-6-phosphate
aminotransferase [isomerizing] GlnS or B3729 SW:GlnS_ECOLI
(P17169) (608 aa) fasta scores: E(): 1e-90, 45.71% id in
630 aa"
/codon_start=1
/transl_table=11
/product="glucosamine--fructose-6-phosphate
aminotransferase [isomerizing]"
/protein_id="CAE50229.1"
/db_xref="GI:38200534"
/translation="MCGIVGFVQRTSVDPDLYALDVLEGLRRLERYGDSAGVAV
ADGAVSFRKAKQVQALQLEQLETSMPQSCIGIHTMTWATGPTDANAPHVVDGK
LAVANGIINFPALGKSELGFGNIFVSEFDEVAATLLGHIENNENAKDLTRAMQLT
CORLEGAFTLLAIHAETPRDIVAARNSPIVIGEGFELGSDVSGFIDTKNAVEN
DNDQIVITITADGHITIDPQGNHAGKGFVVEBMAQAAEKGFYEFMEKEIHEOPAAVR
DTLMGRFDESGLITDLRLIDESTLRSIDKIIVYAGCTAAVAGVAYALIEHMKRITP
EVELAHFRYRDPVINEKTLVITLSQSGETMDILMAVRAHEQAKVIAICNTGSSIT
PRESDACLYTHAGPEIAVASTKAFLAQITATYLLGLAQLRGNMFADENVAVGLAE
TIPDKVAVLDGVDDQVKTALQDMKDATSVLFLGRHYGFVALLEGALKLELAVLAHE
GPAKGLKHGPIILIEEGOVFVVPSPGRSDSLHAKVNSNIOEPRARGAITVIABE
GDPAVEAYAHNIIIRIPAPILMQPLATVLPQIFACGVAAKAKGFDVQPNRLAKSVTV
E"
complement(2651..3073)
/gene="glnS"
/locus_tag="DIP1700"
/note="HMPfam hit to PF01380, SIS domain"
complement(3188..3592)
/gene="glnS"
/locus_tag="DIP1700"
/note="HMPfam hit to PF01380, SIS domain"
complement(3905..4480)
/gene="glnS"
/locus_tag="DIP1700"
/note="HMPfam hit to PF00310, Glutamine amidotransferases
class-II"
complement(4466..4483)
/gene="glnS"
/locus_tag="DIP1700"
/note="ScanRegExp hit to P500443, Glutamine
amidotransferases class-II active site."
4574..5032
/locus_tag="DIP1701"
/locus_tag="5032"
/locus_tag="DIP1701"
/note="Similar to Streptomyces aureofaciens
guanylyl-specific ribonuclease S43 precursor RNAse3
SW:RNS3_STRPU (P30289) (141 aa) fasta scores: E():
1.7e-10, 42.14% id in 140 aa"
/codon_start=1
/transl_table=11
/product="putative ribonuclease"
/protein_id="CAE50230.1"
/db_xref="GI:38200535"
/translation="MSGAKSTPYHKKQSLVAALAGLAIAVAGYFGIQPSQDSFEQA
PATSAIRHEDCEVSESLPQVKDITADITAGCFDYPDNDGVNFGNFGHLPKRDNRN
FYREYVETPDLRHRRGERRITGGGSKTSPQMWYTTDDHYSEFCIRPAH"
4574..4705
/locus_tag="DIP1701"
/note="Signal peptide predicted for DIP1701 by SignalP 2.0
```

HMM (Signal peptide probability 1.000) with cleavage site probability 0.513 between residues 44 and 45:  
 signal-peptide site"  
 misc\_feature  
 4616..4675  
 /locus\_tag="DIP1701"  
 /note="1 probable transmembrane helix predicted for DIP1701 by TMHMM2.0"  
 misc\_feature  
 4754..5026  
 /locus\_tag="DIP1701"  
 /note="TMHMM2am hit to PF00545, ribonuclease."  
 gene  
 5079..5261  
 /locus\_tag="DIP1702"  
 5079..5261  
 /locus\_tag="DIP1702"  
 /note="No significant database matches"  
 /codon\_start=1  
 CDS

Query Match 90.0%; Score 18; DB 1; Length 34895;  
 Best Local Similarity 100.0%; Pred. No. 5e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GAAGTGGGCAATCGTG 20  
 |||||  
 Db 348977 GAAGTGGGCAATCGTG 348994

RESULT 8  
 AC135920/c 143966 bp DNA linear PIN 01-AUG-2004  
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 5 clone  
 OSUNBa0015G13, complete sequence.  
 AC135920  
 AC135920.2 GI:44151402  
 HTG.  
 Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzoideae; Oryza.  
 1 (bases 1 to 143966)  
 Chow, T.-Y., Hsiang, Y.-I. C., Chen, C.-S., Chen, H.-H., Liu, S.-M.,  
 Chao, Y.-T., Chang, S.-J., Chen, H.-C., Chen, S.-K., Chen, T.-R.,  
 Chao, Y.-L., Cheng, C.-H., Chung, C.-I., Han, S.-Y., Hsiao, S.-H.,  
 Hsiung, J.-N., Hsu, C.-H., Huang, J.-J., Kau, P.-I., Lee, M.-C.,  
 Leu, H.-L., Li, Y.-F., Lin, S.-J., Lin, Y.-C., Wu, S.-W., Yu, C.-Y.,  
 Yu, S.-W., Wu, H.-P. and Shaw, J.-F.  
 Oryza sativa BAC OSUNBa0015G13 genomic sequence  
 Unpublished  
 2 (bases 1 to 143966)  
 Chow, T.-Y. and Hsiang, Y.-I. C.  
 Direct Submission  
 Submitted (25-OCT-2002) Institute of Botany, Academia Sinica, 128,  
 Section 2, Academia Road, Nankang, Taipei 11529, Taiwan  
 3 (bases 1 to 143966)  
 Chow, T.-Y.  
 Direct Submission  
 Submitted (27-FEB-2004) Institute of Botany, Academia Sinica, 128,  
 Section 2, Academia Road, Nankang, Taipei 11529, Taiwan  
 4 (bases 1 to 143966)  
 Chow, T.-Y. and Hsiang, Y.-I. C.  
 Direct Submission  
 Submitted (01-AUG-2004) Institute of Botany, Academia Sinica, 128,  
 Section 2, Academia Road, Nankang, Taipei 11529, Taiwan  
 On Feb 27, 2004 this sequence version replaced gi:24371315.  
 Genes were predicated from the integrated results of the following:  
 BLASTN.0, BLASTX.0, GENSCAN (Chris Burge,  
 http://genes.mit.edu/GENSCAN.html), Egenesh  
 (http://www.softberry.com/), glimmer  
 (http://www.tigr.org/softlab/glimmer/glimmer.html), TWINSKAN  
 (http://genes.cs.wustl.edu/) and Genesplinter  
 (http://www.tigr.org/tdb/Genesplinter/index.shtml). The sequence was  
 searched against the Swiss-Prot+TrEMBL protein database, the NCBI  
 Plant EST database, the TIGR Rice Gene Index and the rice

full-length cDNA database (KOME,  
 http://cdna01.dna.affrc.go.jp/cDNA/). Annotated genes are named to  
 indicate the level of evidence for their annotation. Genes with  
 similarity to other proteins are named after the database hits.  
 Genes without significant peptide similarity but with EST  
 similarity are named as unknown proteins. Genes without protein or  
 EST similarity, that are predicted by more than two gene prediction  
 programs over most of their length are annotated as hypothetical  
 proteins. This clone overlaps with R0708D12 (accession # AC130732)  
 and OSUNB00115F21 (accession # AC135918).  
 Location/Qualifiers  
 1..143966  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultiivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="5"  
 /clone="OSUNBa0015G13"  
 3206..5357  
 /gene="OSUNBa0015G13.1"  
 /note="hypothetical protein"  
 join(3206..3386,3466..3731,3918..3959,4266..4449,  
 4684..4883,4977..5357)  
 /gene="OSUNBa0015G13.1"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAT85107.1"  
 /db\_xref="GI:50878332"  
 /translation="MSSRYGVSEGDVHVMNEMLTRDEATHMLLEDFELMESQRRFD  
 DKLRLLQMFQVKEKEMPTKSTGPSATTKAVANGTPRPERVFPPTSLFAPSII  
 SAARATLPSIETAEAGDMQVEEEDTLADICAKVELKORADSLFVELKARPP  
 PTKRSTKTPSIVAQKLLLETNORTLITRSPTEBKVAGLKDKIRLEDVDFMKIL  
 GLHKEVVPKSLIKNMGRAVAVLSIKAWYDERMTLSTKTKIKENKIVVHNIT  
 ILVDPGKNDTLGERVANYMQLQTHMDPEGTGNHRLKESKNSIKRGALGTQJG  
 RMALVMFQSWPKQALAAAYIEQQAALGILKLESKKAASATTAATVAGLGTLL  
 DILLYPSFSFYCTETITTFYSVCIL"  
 13801..15647  
 /gene="OSUNBa0015G13.2"  
 /note="putative metallothionein"  
 join(13801..13850,13996..14043,15548..15647)  
 /gene="OSUNBa0015G13.2"  
 /note="similar to rice EST CF337510"  
 /codon\_start=1  
 /product="putative metallothionein"  
 /protein\_id="AAT85108.1"  
 /db\_xref="GI:50878333"  
 /db\_xref="UniProt/TrEMBL:O22376"  
 /db\_xref="UniProt/TrEMBL:O22376"  
 /translation="MSDKCGNCDACDSQCVKGTSGYGVIVDAEKSHFEMAEVGYE  
 ENDKCKCTTGSCAGCNCGK"  
 complement(125750..26234)  
 /gene="OSUNBa0015G13.3"  
 /note="unknown protein"  
 complement(join(25750..26112,26172..26234))  
 /gene="OSUNBa0015G13.3"  
 /note="similar to rice EST AK107744"  
 /codon\_start=1  
 /product="unknown protein"  
 /protein\_id="AAT85109.1"  
 /db\_xref="GI:50878334"  
 /db\_xref="UniProt/TrEMBL:O9AW66"  
 /translation="MARLIGIGVAAAVVAVATVPAAHHPPPRHLLLVVRQGA  
 VLSGGGRRRAAPAGAGSEADDEPRLDGLGREGGAVQOARPLLPELRVAAAGESL  
 PRLRLRYPLPQELTMMVMVTSLSIQIPFIRPIH"  
 complement(36352..36954)  
 /gene="OSUNBa0015G13.4"  
 /note="hypothetical protein"  
 complement(36352..36954)  
 /gene="OSUNBa0015G13.4"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAT85110.1"  
 /db\_xref="GI:50878335"  
 /db\_xref="UniProt/TrEMBL:Q9SYMS"

CDS

gene

gene

CDS

gene

CDS

gene

FEATURES  
source

```

COMPLEMENT (join(88454. .90352,.95721. .96066,.96167. .96189)
/genes="OSJNB.0015G13.9"
/notes="similar to rice EST AK069504"
/codon_start=1
/product="unknown protein"
/protein_id="AA185115.1"
/db_xref="GI:50878340"
/db_xref="UniProt/TxEMBL:Q9AW18"
/translation="MAPSFIREYAAASMRGTAAASRHPALSAAGDFPMPPIRRFRMWYD")
Query Match      87.0%; Score 17.4; DB 8; Length 143966;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
By
1 TCGAAGGTGGGCAATCGT 19
|||||
Db 91684 TCGAAGGTGGGCAATCAT 91666

RESULT 9
AC122848 204254 bp DNA linear ROD 01-JAN-2004
LOCUS AC122848
DEFINITION Mus musculus BAC clone RP23-16K13 from chromosome 18, complete
sequence.
ACCESSION AC122848
VERSION AC122848.3 GI:38564454
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 204254)
TITLE Tomlinson,C., Cotton,M., Bielicki,L. and Haskenson,W.
JOURNAL The sequence of Mus musculus BAC clone RP23-16K13
REFERENCE
AUTHORS 2 (bases 1 to 204254)
TITLE Wilson,R.
JOURNAL Sequencing of Mus musculus
REFERENCE
AUTHORS 3 (bases 1 to 204254)
TITLE McPherson,J.D. and Waterston,R.H.
JOURNAL Direct Submission
REFERENCE
AUTHORS 4 (bases 1 to 204254)
TITLE McPherson,J.D. and Waterston,R.H.
JOURNAL Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS 5 (bases 1 to 204254)
TITLE Wilson,R.
JOURNAL Submitted (28-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS 6 (bases 1 to 204254)
TITLE Wilson,R.
JOURNAL Direct Submission
REFERENCE
AUTHORS Submitted (01-JAN-2004) Department of Genetics, Washington
TITLE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL Un Nov 28, 2003 this sequence version replaced gi:22657947.
COMMENT
-----
Center: Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
-----
Summary Statistics
Center project name: M_BA016K13
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

```

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The RPc1-23 BAC library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC132427 and AC125110.

## FEATURES

## SOURCE

1.204254

Location/Qualifiers

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="18"

/map="18"

/clone="RP23-16K13"

/clone\_lib="RPc1-23"

247.376

/rpt\_family="Alu"

756.808

/rpt\_family="MIR"

1826.2049

/rpt\_family="MER1\_type"

2425.2601

/rpt\_family="B2"

2758.2898

/rpt\_family="Alu"

3672.3697

/rpt\_family="ERVK"

4709.4870

/rpt\_family="L1"

4972.5076

/rpt\_family="Alu"

5144.5317

/rpt\_family="B2"

5703.6147

/rpt\_family="MALR"

6489.6887

/rpt\_family="MALR"

8817.8898

/rpt\_family="MALR"

7826.8220

/rpt\_family="MALR"

9632.9778

/rpt\_family="Alu"

11045.11290

/rpt\_family="L1"

12803.12897

/rpt\_family="B4"

13109.13326

/rpt\_family="L1"

13358.13470

/rpt\_family="L1"

13752.13834

/rpt\_family="Alu"

repeat\_region

14694.14846

/rpt\_family="L2"

repeat\_region

14867.14889

/rpt\_family="B4"

repeat\_region

14890.15026

/rpt\_family="Alu"

repeat\_region

16119.16197

/rpt\_family="B4"

repeat\_region

16280.16412

/rpt\_family="Alu"

repeat\_region

17016.17120

/rpt\_family="Alu"

repeat\_region

17481.17686

/rpt\_family="MER1\_type"

repeat\_region

19443.19381

/rpt\_family="MALR"

repeat\_region

20432.20549

/rpt\_family="B4"

repeat\_region

21468.21577

/rpt\_family="MIR"

repeat\_region

21738.21817

/rpt\_family="L2"

repeat\_region

21808.21970

/rpt\_family="MIR"

repeat\_region

21975.22069

/rpt\_family="MALR"

repeat\_region

21991.22098

/rpt\_family="Alu"

repeat\_region

22194.22255

/rpt\_family="ERV1"

repeat\_region

22373.22471

/rpt\_family="ID"

repeat\_region

23394.23559

/rpt\_family="MER1\_type"

repeat\_region

23905.24003

/rpt\_family="MIR"

repeat\_region

24667.24777

/rpt\_family="L1"

repeat\_region

24778.24827

/rpt\_family="B4"

repeat\_region

24857.25075

/rpt\_family="L1"

repeat\_region

25108.25200

/rpt\_family="MER1\_type"

repeat\_region

25824.26052

/rpt\_family="B2"

repeat\_region

26339.26390

/rpt\_family="L2"

repeat\_region

26595.26667

/rpt\_family="MIR"

repeat\_region

26852.26896

/rpt\_family="MER1\_type"

repeat\_region

27760.27818

/rpt\_family="L2"

repeat\_region

28365.28438

/rpt\_family="L1"

repeat\_region

30127.30197

/rpt\_family="ERV1"

repeat\_region

31523.31724

/rpt\_family="MIR"

repeat\_region

31885.32031

/rpt\_family="Alu"

repeat\_region

33585.33778

/rpt\_family="MIR"

repeat\_region

33809.34232

/rpt\_family="MALR"

repeat\_region

34452.34491

/rpt\_family="L1"

repeat\_region

34483.34655

/rpt\_family="L1"

repeat\_region

34822.35154

/rpt\_family="L1"

repeat\_region

35740.35841



```

repeat_region      /rpt family="MALR"
                   36342. .36552

Query Match      87.0%; Score 17.4; DB 10; Length 204254;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      2 CGAAGTGGCGCAATCGT 20
      |||||
      114915 CCAAGTGGCGCAATCGT 114913

RESULT 10
AC146138      176823 bp      DNA      linear      PRI 19-MAY-2004
LOCUS      Pan troglodytes BAC clone RP43-3D14 from 7, complete sequence.
ACCESSION      AC146138
VERSION      AC146138.2 GI:46064773
KEYWORDS      HTG.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM      Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
1 (bases 1 to 176823)
Kozlowicz, A. and Isak, A.
The sequence of Pan troglodytes BAC clone RP43-3D14
Unpublished (2001)
2 (bases 1 to 176823)
Wilson, R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 176823)
Wilson, R.K.
Direct Submission
Submitted (02-APR-2004) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 176823)
Wilson, R.K.
Direct Submission
Submitted (19-MAY-2004) Washington University School of Medicine,
Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO
63108, USA
On Apr 2, 2004 this sequence version replaced gi:33387092.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics -----
Center project name: C_PT003D14

NOTICE:
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RP43-BAC library has been constructed by Chung-Li Shu. DNA
was isolated from white blood cells obtained from a male chimpanzee
(Pan troglodytes, 'Clint', Yerkes #C0471; Birthdate: 6-6-80). The

```

```

clone and detailed information can be obtained from Resgen
(http://www.resgen.com) or Pieter de Jong and co-workers at
http://www.bacpac.chori.org.

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1. 176823
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/map="7"
/clone="RP43-3D14"
/clone_1fb="RPCI-43"
73054..73089
/notes="Sequence derived from one plasmid subclone."
108198..108226
/notes="Sequence derived from one plasmid subclone."

ORIGIN
Query Match      85.0%; Score 17; DB 9; Length 176823;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2 CGAAGTGGCGCAATCG 18
      |||||
      169865 CCAAGTGGCGCAATCG 169881

RESULT 11
AC146015      178455 bp      DNA      linear      HTG 24-APR-2004
LOCUS      Pan troglodytes chromosome 7 clone RP43-129D8, WORKING DRAFT
DEFINITION      SEQUENCE, 10 unordered pieces.
ACCESSION      AC146015
VERSION      AC146015.3 GI:46559526
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ACTIVEFN.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM      Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
1 (bases 1 to 178455)
Wilson, R.K.
The sequence of Pan troglodytes clone
Unpublished
2 (bases 1 to 178455)
Wilson, R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 178455)
Wilson, R.K.
Direct Submission
Submitted (24-APR-2004) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
On Apr 24, 2004 this sequence version replaced gi:35071906.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: C_PT129D08
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 173793 bases at least Q40

```



Consensus quality: 174974 bases at least Q30  
Consensus quality: 175572 bases at least Q20

NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1  
1242: contig of 1242 bp in length  
1243  
1342: gap of unknown length  
1343  
2915: contig of 1573 bp in length  
2916  
3015: gap of unknown length  
3016  
4527: contig of 1512 bp in length  
4528  
4627: gap of unknown length  
4628  
7912: contig of 3285 bp in length  
7913  
8012: gap of unknown length  
8013  
20454: contig of 12442 bp in length  
20455  
20554: gap of unknown length  
27431: contig of 6877 bp in length  
27432  
27531: gap of unknown length  
27532  
74998: contig of 47467 bp in length  
74999  
75098: gap of unknown length  
75099  
106492: contig of 31394 bp in length  
106493  
147728: contig of 41136 bp in length  
147729  
147829: gap of unknown length  
147829  
178455: contig of 30627 bp in length.

## FEATURES

## source

/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="7"  
/clone="RP43-129D8"  
1..1242  
/note="assembly\_name:Contig42"  
1343..2915  
/note="assembly\_name:Contig45"  
3016..4527  
/note="assembly\_name:Contig48  
clone end:SP6  
vector\_side:right"  
4628..7912  
/note="assembly\_name:Contig49"  
8013..20454  
/note="assembly\_name:Contig50"  
20555..27431  
/note="assembly\_name:Contig51"  
27532..74998  
/note="assembly\_name:Contig52"  
75099..106492  
/note="assembly\_name:Contig53"  
106593..147728  
/note="assembly\_name:Contig54"  
147829..178455  
/note="assembly\_name:Contig55"

## ORIGIN

Query Match 85.0%; Score 17; DB 2; Length 178455;  
Best Local Similarity 100.0%; Pred. NO. 1.6e+0; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0;

Qy 2 CGAAGTGGCAATCG 18  
|||||

Db 85883 CGAAGTGGCAATCG 85899

RESULT 12  
CEY102ASD  
LOCUS CEY102ASD 25454 bp DNA linear INV 12-OCT-2004

## DEFINITION

Caenorhabditis elegans YAC Y102ASD, complete sequence.

## ACCESSION

AL031628  
AL031628.2 GI:6425213

## VERSION

HTG.

## KEYWORDS

Caenorhabditis elegans

## SOURCE

Caenorhabditis elegans

## ORGANISM

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditidae; Rhabditidae; Pelodermidae; Caenorhabditis.

## REFERENCE

1

## AUTHORS

none.

## TITLE

Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)

## JOURNAL

99069613

## MEDLINE

9851916

## REMARK

The C. elegans Sequencing Consortium.

## REFERENCE

2 (bases 1 to 25454)

## AUTHORS

Gardner,A.E.

## TITLE

Direct Submission

## JOURNAL

Submitted (22-SEP-1998) Nematode Sequencing Project, Sanger  
Institute, Hinxton, Cambridge CB10 1SA, England and Department of  
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
jess@sanger.ac.uk or rwnematode.wustl.edu B-mail: worm@sanger.ac.uk  
On Nov 15, 1999 this sequence version replaced gi:3646777.  
Coding sequences below are predicted from computer analysis, using  
predictions from GeneFINDER (P. Green, U. Washington), and other  
available information.  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.  
For a graphical representation of this sequence and its analysis  
see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?  
name=Y102ASD  
For a graphical representation of this sequence and its analysis  
see: - http:  
5S RNA tandem repeat, EMBL accession X16224. (Nelson & Honda Gene  
(1985) 38:245-251). It starts in cosmid CEK218. Approximately  
90-100 copies of the 1kb repeat are thought to exist. Due to  
difficulties in accurately sequencing the tandem array only 12  
copies are represented here. At least 2 forced joins have been made  
and this is one region where sequence has been derived from a  
single clone  
IMPORTANT: This sequence is not the entire insert of clone Y102ASD.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone F40D4 is at 25351 in this sequence. The  
true right end of clone ZK218 is at 104 in this sequence. The start  
of this sequence (1..104) overlaps with the end of sequence 282085.  
The end of this sequence (25351..25454) overlaps with the start of  
sequence 281536.

## FEATURES

## source

Location/Qualifiers  
1..25454  
/organism="Caenorhabditis elegans"  
/mol\_type="genomic DNA"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="V"  
/clone="Y102ASD"  
18800..18821  
/gene="Y102ASD.4"  
18800..18821  
/gene="Y102ASD.4"  
/product="Hypothetical RNA transcript Y102ASD.4"  
/standard\_name="Y102ASD.4"  
complement(join(20322..20673,20849..21515,21771..22002))  
/gene="Y102ASD.1"

```

CDS                                     complement {join(20322..20673,20849..21515,21771..22002)}
                                         /gene="Y102ASD.1"
                                         /standard_name="Y102ASD.1"
                                         /codon_start=1
                                         /product="Hypothetical protein Y102ASD.1"
                                         /protein_id="CAA20973.1"
                                         /db_xref="gi:1806666"
                                         /db_xref="UniProt/TREMBL/Q9XK64"
                                         /translation="MAEIDAVDTGVTVNIAQGTATLTAGOVVFAKVRDVETRVDC
OILQEDMIVILTTFKMKNTEKERCESTGDLDEPRSGPLEVIQIPVKDGIGVACPTT
TCMTITPLVSTLKAKMSSDDYIKNIYTYSAYEVNQTKGKKNDVEDIAEFLANERRYY
RSNGVFPSRLVDYIRNLVSITYGNAPLHKVAVNRNGKLASVQTINSAELTYLDADI
HKLDQYLALVNKAALKLFTEKGPKVKEVAVWVGQNSINKFVNHGGYLPACGRFVH
PALNLLACVANSFDQTRKEKRSMKKSNTGGCHLPRTWNRSDEKTLRKRRKISLS
WLOHAQENQNAVCPVDRLSLKSVIPELCATVMESQSKAFNMHIIGDKYTDVPTEL
GHRRPKGNQMKEAHPHARSTGN"

ORIGIN
Query Match      84.0%; Score 16.8; DB 3; Length 25454;
Best Local Similarity 90.0%; Pred. No. 2.le+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      1 TCGAAGTGCGCAATCGTG 20
        ||| ||||| ||||| |||
Db      12033 TCGCAGGTGCGCAATCTCG 12052

RESULT 13
LOCUS       AL669827             108424 bp    DNA          linear   ROD 27-APR-2002
DEFINITION  Mouse DNA sequence from clone RP23-350D17 on chromosome 11,
complete sequence.
ACCESSION   AL669827
KEYWORDS    AL669827.5 GI:20339131
SOURCE      HTG.
ORGANISM    Mus musculus (house mouse)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 108424)
            Oliver,K.
DIRECT SUBMISSION
Submitted (27-Apr-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SN, UK. E-mail enquiries: clonequest@sanger.ac.uk
hungquery@sanger.ac.uk Clone request: clonequest@sanger.ac.uk
On Apr 29, 2002 this sequence version replaced gi:20218615.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Emi, EMBL; Sw:,
SWISSPOT; Tr:, TREMBL; Mp:, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/RP23-350D17.htm
from the RPEC1-23 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6.
Location/Qualifiers
1..108424
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
```

```

ORIGIN
Query Match      84.0%; Score 16.8; DB 10; Length 108424;
Beet Local Similarity 90.0%; Pred. No. 2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY          1 TCGAAGGTGGGCAATCGTG 20
           |||
DB          62993 TCCGAGTGGGCACAAAGCTGTG 63012

RESULT 14
LOCUS       AC136469              147180 bp    DNA             linear   PRI 05-DEC-2002
DEFINITION  Homo sapiens chromosome 19 clone RP11-886P16, complete sequence.
ACCESSION   AC136469
VERSION     AC136469.2   GI:26050948
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 147180)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            Direct Submission
            Unpublished
            2 (bases 1 to 147180)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            Direct Submission
            Submitted (02-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            3 (bases 1 to 147180)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            Direct Submission
            Submitted (05-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            On Dec 5, 2002 this sequence version replaced gi:24476050.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            www.sngc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.4.
FEATURES             source
     source          1..147180
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="19"
                        /clone="RP11-886P16"

ORIGIN
Query Match      84.0%; Score 16.8; DB 9; Length 147180;
Beet Local Similarity 90.0%; Pred. No. 2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY          1 TCGAAGGTGGGCAATCGTG 20
           |||
DB          137860 TCCGAAGTGGGCACATCGTG 137879

RESULT 15
LOCUS       AC136470              156942 bp    DNA             linear   HTG 02-NOV-2002
DEFINITION  Homo sapiens chromosome 19 clone RPCI-13_947L3, *** SEQUENCING IN
PROGRESS ***, 3 ordered pieces.
ACCESSION   AC136470
VERSION     AC136470.1   GI:24476051
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEPIN.
            Homo sapiens (human)
SOURCE      Homo sapiens (human)

```

```

ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL       1 (bases 1 to 156942)
REFERENCE     DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS       Direct Submission
TITLE         Unpublished
JOURNAL       2 (bases 1 to 156942)
REFERENCE     DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS       Direct Submission
TITLE         Submitted (02-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL       Drive, Walnut Creek, CA 94598, USA
COMMENT       * NOTE: This is a 'working draft' sequence. It currently
              * consists of 3 contigs. Gaps between the contigs
              * are represented as runs of N. The order of the pieces
              * is believed to be correct as given, however the sizes
              * of the gaps between them are based on estimates that have
              * provided by the submittor.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.
              *
              * 1 10649: contig of 10649 bp in length
              * 10650 10749: gap of unknown length
              * 10750 138483: contig of 127734 bp in length
              * 138484 138583: gap of unknown length
              * 138584 156942: contig of 18359 bp in length.
              -----Genome Center
              Center: Joint Genome Institute
              Center Code: JGI
              Web site: http://www.jgi.doe.gov
              -----
              Project Information
              Center Project Name: 3368425
              Center clone name: RPCI-13_947L3
              -----
              Summary Statistics
              Consensus quality: 155815 bases at least Q40
              Consensus quality: 155971 bases at least Q30
              Consensus quality: 156062 bases at least Q20
              Estimated insert size: 155000; agarose-fp estimation
              Estimated insert size: 156296; sum-of-contigs

estimation
estimation
estimation
Quality coverage: 1 in Q20 bases; agarose-fp
Quality coverage: 0.99 in Q20 bases; sum-of-contigs
estimation.
FEATURES
source
Location/Qualifiers
1..156942
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RPCI-13_947L3"
ORIGIN
Query Match 84.0%; Score 16.8; DB 2; Length 156942;
Best Local Similarity 90.0%; Pred. No. 2e+03; 2; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TCGAAGTGGCGCAATCGTG 20
DB 124563 TCGAAGTGGCGCGATCGTG 124582

```

Search completed: November 7, 2005, 06:41:59  
 Job time : 795.957 secs

This Page Blank Inspire

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21 ; Search time 230.638 Seconds  
(without alignments)  
513.336 Million cell updates/sec

Title: US-10-777-131A-3

Perfect score: 20

Sequence: 1 tcgaagtggtggcaatcgtg 20

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001s:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	ABZ79929	Abz79929 Mycobacter
2	20	100.0	1144	ABZ79953	Abz79953 Mycobacter
3	20	100.0	103765	AA199683_43	Continuation (44 o
4	20	100.0	110000	AA199682_43	Continuation (44 o
5	16	80.0	511	ADN13107	Adn13107 Human pro
6	16	80.0	2267	AA197922	Aa197922 Human neu
7	16	80.0	2267	AA198077	Aa198077 Human neu
8	15.8	79.0	602	AC119646	Ac119646 DNA clone
9	15.8	79.0	622	AC119645	Ac119645 DNA clone
10	15.8	79.0	663	ABZ53602	Abz53602 Human CDN
11	15.8	79.0	714	AC119647	Ac119647 DNA clone
12	15.8	79.0	837	ABZ53601	Abz53601 Human CDN
13	15.8	79.0	950	ABZ12053	Abz12053 Human pol
14	15.8	79.0	950	ADMA4571	Adma4571 Novel hum
15	15.8	79.0	1104	ADFS3637	Adfs3637 Murine DN
16	15.8	79.0	1165	AAZ23915	Aaz23915 Human dua
17	15.8	79.0	1165	ADN75961	Adn75961 Human sig
18	15.8	79.0	1165	ADN76019	Adn76019 Human DSP
19	15.8	79.0	1262	AAZ09495	Aaz09495 Human SGP
20	15.8	79.0	1422	AAZ29556	Aaz29556 M.catarrh

C	21	15.8	79.0	1891	10	ADB63795	Adb63795 Human CDN
C	22	15.8	79.0	2241	5	AAS68586	Aas68586 DNA encod
C	23	15.8	79.0	2395	4	AB113174	Ab113174 Drosophill
C	24	15.8	79.0	3252	4	AB106532	Ab106532 Drosophill
C	25	15.8	79.0	3520	13	ADS49321	Ads49321 Bacterial
C	26	15.8	79.0	4596	4	AB102235	Ab102235 Drosophill
C	27	15.8	79.0	7149	4	AB102234	Ab102234 Drosophill
C	28	15.8	79.0	95001	12	ADH56439	Adh56439 Human hyp
C	29	15.8	79.0	110000	12	AD078172_1	Continuation (2 of
C	30	15.8	79.0	117328	13	ABD32886	Abd32886 Mouse can
C	31	15.8	79.0	121162	3	AAc66548	Aac66548 Human kin
C	32	15.8	79.0	145606	12	ADP44617	Adp44617 Human alp
C	33	15.8	79.0	349980	6	ABQ81842	Abq81842 Bifidobac
C	34	15.8	79.0	349980	13	ADT05737	Adt05737 Haemophil
C	35	15.4	77.0	360	4	AA186669	Aa186669 Human pol
C	36	15.4	77.0	822	12	ADJ42411	Adj42411 Plant CDN
C	37	15.4	77.0	825	13	ADT45067	Adt45067 Bacterial
C	38	15.4	77.0	1191	11	ADM92655	Adm92655 SNP-conta
C	39	15.4	77.0	1928	2	AAQ32139	Aaq32139 Bauri pro
C	40	15.4	77.0	3337	4	AB114924	Ab114924 Drosophill
C	41	15.4	77.0	4343	13	ADR07326	Adr07326 Full leng
C	42	15.4	77.0	23075	5	ABA19112	Abal19112 Human mer
C	43	15.4	77.0	49939	9	ADB16928	Adb16928 Human DYX
C	44	15.4	77.0	110000	6	ABQ69245_09	Continuation (10 o
C	45	15.4	77.0	110000	6	ABQ67197_08	Continuation (9 of

## ALIGNMENTS

RESULT 1	ABZ79929	standard; DNA; 20 BP.
ID	ABZ79929	
XX	ABZ79929;	
XX	19-MAY-2003 (first entry)	
XX	Mycobacterium tuberculosis Rv3908-1-1 PCR primer SEQ ID NO:3.	
XX	Mycobacterium tuberculosis; mutT2; alka; ogt; Rv3908; mutY; Rv3909;	
XX	KW detection; multidrug resistance; multiple drug resistance; MDR;	
XX	KW infection; PCR primer; ss.	
XX	Mycobacterium tuberculosis.	
OS	Synthetic.	
XX	WO2003016562-A2.	
XX	27-FEB-2003.	
XX	14-AUG-2002; 2002WO-EP009679.	
XX	14-AUG-2001; 2001US-0311824P.	
XX	21-AUG-2001; 2001US-0313523P.	
XX	(INSP ) INST PASTEUR.	
XX	Gicquel B;	
XX	WPI; 2003-256711/25.	
XX	Predicting the epidemic character of a Mycobacterium tuberculosis isolate	
XX	and/or the acquisition of multiple drug resistance (MDR) by the isolate	
XX	by detecting an alteration in the DNA repair system of the isolate.	
XX	Claim 32; Page 16; 83pp; English.	
XX	The present invention describes a method for predicting the epidemic	
XX	character of a Mycobacterium tuberculosis isolate and/or a selective	
XX	advantage to be maintained in the host and/or the acquisition of multiple	
XX	drug resistance (MDR) by the isolate comprising detecting an alteration	
XX	in the DNA repair system of the isolate. Also described: (1) detecting a	

CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a PCR primer for M. tuberculosis Rv3908,  
CC which is used in the exemplification of the present invention  
XX  
SQ Sequence 20 BP; 5 A; 3 C; 8 G; 4 T; 0 U; 0 Other;  
XX  
Query Match 100.0%; Score 20; DB 8; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCGAAGGTGGGCAATCGTG 20  
DB 1 TCGAAGGTGGGCAATCGTG 20  
XX  
RESULT 2  
ABZ79953  
ID ABZ79953 standard; DNA; 1144 BP.  
XX  
AC ABZ79953;  
XX  
DT 19-MAY-2003 (first entry)  
XX  
DE Mycobacterium tuberculosis Rv3908 nucleotide sequence SEQ ID NO:20.  
XX  
KW Mycobacterium tuberculosis; multT2; alrA; ogt; Rv3908; mutY; Rv3909;  
KW detection; multidrug resistance; multiple drug resistance; MDR;  
KW infection; gene; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO2003016562-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-EP009679.  
XX  
PR 14-AUG-2001; 2001US-0311824P.  
PR 21-AUG-2001; 2001US-0313523P.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI G1cquel B;  
XX  
DR WPI, 2003-256711/25.  
XX  
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate  
PT by detecting an alteration in the DNA repair system of the isolate.  
XX  
PS Disclosure; Fig 5A; 83pp; English.  
XX  
XX The present invention describes a method for predicting the epidemic  
CC character of a Mycobacterium tuberculosis isolate and/or a selective  
CC advantage to be maintained in the host and/or the acquisition of multiple  
CC drug resistance (MDR) by the isolate comprising detecting an alteration  
CC in the DNA repair system of the isolate. Also described: (1) detecting a  
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a M. tuberculosis Rv3908 nucleotide

CC sequence, which is used in the exemplification of the present invention  
XX  
SQ Sequence 1144 BP; 194 A; 405 C; 379 G; 166 T; 0 U; 0 Other;  
XX  
Query Match 100.0%; Score 20; DB 8; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCGAAGGTGGGCAATCGTG 20  
DB 7 TCGAAGGTGGGCAATCGTG 26  
XX  
RESULT 3  
AA199683 43  
Continuation (44 of 44) of AA199683 from base 4300001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683  
WP Fragment Name Begin End  
WP AA199683\_00 1 110000  
WP AA199683\_01 100001 210000  
WP AA199683\_02 200001 310000  
WP AA199683\_03 300001 410000  
WP AA199683\_04 400001 510000  
WP AA199683\_05 500001 610000  
WP AA199683\_06 600001 710000  
WP AA199683\_07 700001 810000  
WP AA199683\_08 800001 910000  
WP AA199683\_09 900001 1010000  
WP AA199683\_10 1000001 1110000  
WP AA199683\_11 1100001 1210000  
WP AA199683\_12 1200001 1310000  
WP AA199683\_13 1300001 1410000  
WP AA199683\_14 1400001 1510000  
WP AA199683\_15 1500001 1610000  
WP AA199683\_16 1600001 1710000  
WP AA199683\_17 1700001 1810000  
WP AA199683\_18 1800001 1910000  
WP AA199683\_19 1900001 2010000  
WP AA199683\_20 2000001 2110000  
WP AA199683\_21 2100001 2210000  
WP AA199683\_22 2200001 2310000  
WP AA199683\_23 2300001 2410000  
WP AA199683\_24 2400001 2510000  
WP AA199683\_25 2500001 2610000  
WP AA199683\_26 2600001 2710000  
WP AA199683\_27 2700001 2810000  
WP AA199683\_28 2800001 2910000  
WP AA199683\_29 2900001 3010000  
WP AA199683\_30 3000001 3110000  
WP AA199683\_31 3100001 3210000  
WP AA199683\_32 3200001 3310000  
WP AA199683\_33 3300001 3410000  
WP AA199683\_34 3400001 3510000  
WP AA199683\_35 3500001 3610000  
WP AA199683\_36 3600001 3710000  
WP AA199683\_37 3700001 3810000  
WP AA199683\_38 3800001 3910000  
WP AA199683\_39 3900001 4010000  
WP AA199683\_40 4000001 4110000  
WP AA199683\_41 4100001 4210000  
WP AA199683\_42 4200001 4310000  
WP AA199683\_43 4300001 4403765  
XX  
Query Match 100.0%; Score 20; DB 4; Length 103765;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCGAAGGTGGGCAATCGTG 20  
DB 85490 TCGAAGGTGGGCAATCGTG 85509  
XX  
RESULT 4

AA199682.43  
Continuation (44 of 45) of AA199682 from base 4300001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682  
WP Fragment Name Begin End  
WP AA199682\_00 1 110000  
WP AA199682\_01 100001 210000  
WP AA199682\_02 200001 310000  
WP AA199682\_03 300001 410000  
WP AA199682\_04 400001 510000  
WP AA199682\_05 500001 610000  
WP AA199682\_06 600001 710000  
WP AA199682\_07 700001 810000  
WP AA199682\_08 800001 910000  
WP AA199682\_09 900001 1010000  
WP AA199682\_10 1000001 1110000  
WP AA199682\_11 1100001 1210000  
WP AA199682\_12 1200001 1310000  
WP AA199682\_13 1300001 1410000  
WP AA199682\_14 1400001 1510000  
WP AA199682\_15 1500001 1610000  
WP AA199682\_16 1600001 1710000  
WP AA199682\_17 1700001 1810000  
WP AA199682\_18 1800001 1910000  
WP AA199682\_19 1900001 2010000  
WP AA199682\_20 2000001 2110000  
WP AA199682\_21 2100001 2210000  
WP AA199682\_22 2200001 2310000  
WP AA199682\_23 2300001 2410000  
WP AA199682\_24 2400001 2510000  
WP AA199682\_25 2500001 2610000  
WP AA199682\_26 2600001 2710000  
WP AA199682\_27 2700001 2810000  
WP AA199682\_28 2800001 2910000  
WP AA199682\_29 2900001 3010000  
WP AA199682\_30 3000001 3110000  
WP AA199682\_31 3100001 3210000  
WP AA199682\_32 3200001 3310000  
WP AA199682\_33 3300001 3410000  
WP AA199682\_34 3400001 3510000  
WP AA199682\_35 3500001 3610000  
WP AA199682\_36 3600001 3710000  
WP AA199682\_37 3700001 3810000  
WP AA199682\_38 3800001 3910000  
WP AA199682\_39 3900001 4010000  
WP AA199682\_40 4000001 4110000  
WP AA199682\_41 4100001 4210000  
WP AA199682\_42 4200001 4310000  
WP AA199682\_43 4300001 4410000  
WP AA199682\_44 4400001 4411529

Query Match 100.0%; Score 20; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 3.8; Mismatches 0; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCGAGGTGGCAATCGTG 20  
DB 93252 TCGAGGTGGCAATCGTG 93271

RESULT 5  
ADN13107/c  
ID ADN13107 standard; cDNA; 511 BP.  
XX AC ADN13107;  
XX DT 29-JUN-2004 (first entry)  
XX DE Human prostate/colon/lung/breast cancer-related cDNA 622, SEQ:622.  
XX KW Human; cancer; tumour; prostate cancer; colon cancer; lung cancer;  
KW breast cancer; drug screening; diagnosis; prognosis; prevention;  
KW gene mapping; tissue typing; tissue profiling; cytostatic; gene therapy;  
KW 85.

XX OS Homo sapiens.  
XX PN WO2004039943-A2.  
XX PD 13-MAY-2004.  
XX PF 16-MAY-2003; 2003WO-US015465.  
XX PR 17-MAY-2002; 2002US-0381533P.  
XX PR 04-FEB-2003; 2003US-0445222P.  
XX PA (CHIR ) CHIRON CORP.  
XX PI Scott EM, Lamson G, Kassam A, Zhang G, Sakamoto D, Garcia PD;  
XX DR WPI; 2004-376173/35.  
XX PT New isolated polynucleotides, useful for gene mapping or tissue typing or  
XX PT profiling, as diagnostic reagents, and for preventing or treating cancer,  
XX PT e.g. prostate, colon, or breast cancer.  
XX PS Claim 2; SEQ ID NO 622; 190pp; English.  
XX CC The invention relates to nucleic acids (ADN12486-ADN13970) isolated from  
XX CC human prostate, colon, lung and breast cancer cDNA libraries, and to 57  
XX CC proteins (ADN13971-ADN14027) encoded by a subset of these cDNA sequences  
XX CC (ADN13914-ADN13970). The invention also relates to vectors and host cells  
XX CC comprising a nucleic acid of the invention; a method for the recombinant  
XX CC production of a protein of the invention; an antibody specific for a  
XX CC protein of the invention; a polynucleotide library comprising at least  
XX CC one nucleic acid sequence of the invention; a method for detecting a  
XX CC cancerous cell by PCR or probe hybridisation; inhibiting a cancerous  
XX CC phenotype (particularly aberrant proliferation) of a cell; a method of  
XX CC identifying an agent that modulates the biological activity of a gene  
XX CC product differentially expressed in a cancerous cell compared with a  
XX CC normal cell; and a method of treating a cancer patient using the agent  
XX CC identified. The nucleic acids and polypeptides can be used to diagnose,  
XX CC prognose, treat or prevent cancers such as prostate, colon, lung or  
XX CC breast cancer, and can also be used to screen for drugs for the treatment  
XX CC of cancer. The nucleic acids can also be used for gene mapping, tissue  
XX CC typing and tissue profiling. The present sequence represents a  
XX CC specifically claimed cancer-related cDNA of the invention. Note: The  
XX CC sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 511 BP; 131 A; 115 C; 122 G; 143 T; 0 U; 0 Other;  
XX

Query Match 80.0%; Score 16; DB 12; Length 511;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 5 AGGTGGCAATCGTG 20  
DB 116 AGGTGGCAATCGTG 101

RESULT 6  
AA197922  
ID AA197922 standard; cDNA; 2267 BP.  
XX AC AA197922;  
XX DT 13-NOV-2001 (first entry)  
XX DE Human neuroblastoma expressed polynucleotide SEQ ID NO 3997.  
XX KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; 85.  
XX OS Homo sapiens.  
XX PN WO200166719-A1.

```
XX 13-SEP-2001.
PD 02-MAR-2001; 2001WO-JP001629.
XX 07-MAR-2000; 2000JP-00159195.
PR 07-MAR-2000; 2000JP-00159195.
XX (CHIB-) CHIBA PREFECTURE.
PA (HISM ) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
PI WPI; 2001-565584/63.
XX Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human neuroblastoma,
PT malignancy and susceptibility indicator or tumor marker for anti-cancer
PT agents.
XX Claim 1; Page 2945-2946; 2979pp; Japanese.
XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes
XX Sequence 2267 BP; 755 A; 439 C; 483 G; 590 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16; DB 4; Length 2267;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 CGAAGTGGCGCAATC 17
DB 1970 CGAAGTGGCGCAATC 1985
RESULT 7
AAI98077
ID AAI98077 standard; cDNA; 2267 BP.
XX AAI98077;
XX 04-DEC-2001 (first entry)
XX Human neuroblastoma expressed polynucleotide SEQ ID NO 20.
DE Human neuroblastoma; ss.
XX Human; neuroblastoma; ss.
XX Homo sapiens.
XX WO200166733-A1.
XX 13-SEP-2001.
XX 02-MAR-2001; 2001WO-JP001631.
XX 07-MAR-2000; 2000JP-00159195.
PR 12-MAY-2000; 2000JP-00140387.
XX (CHIB-) CHIBA PREFECTURE.
PA (HISM ) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
PI WPI; 2001-602630/68.
XX Nucleic acids for prognosis of human neuroblastoma comprise nucleic acids
PT expressed by human neuroblastomas.
PS Claim 1; Page 66-67; 159pp; Japanese.
```

```
XX The invention relates to nucleic acids (AAI98058-AAI98161) or their
CC homologues expressed by human neuroblastomas useful for detecting genes
CC expressed by neuroblastoma and for analysing their structure and
CC function. The nucleic acids are useful for the diagnosis and prognosis of
CC neuroblastoma
XX Sequence 2267 BP; 755 A; 439 C; 483 G; 590 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16; DB 4; Length 2267;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 CGAAGTGGCGCAATC 17
DB 1970 CGAAGTGGCGCAATC 1985
RESULT 8
ACLI9646/C
ID ACLI9646 standard; DNA; 602 BP.
XX ACLI9646;
XX 17-OCT-2003 (first entry)
XX DNA clone originating in barley containing SNP encoding sequence #9637.
XX Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
XX gene; ss.
XX Hordeum vulgare.
XX WO2003057877-A1.
XX 17-JUL-2003.
XX 16-DEC-2002; 2002WO-IB005403.
XX 20-DEC-2001; 2001JP-00387059.
PR 20-DEC-2001; 2001JP-00387131.
PR 20-DEC-2001; 2001JP-00403289.
PR 20-DEC-2001; 2001JP-00403300.
PR 27-SEP-2002; 2002JP-00327515.
XX (UNNI-) UNIV JAPAN OKAYAMA.
XX Sato K, Takeda K, Kohara Y;
PI WPI; 2003-587127/55.
XX Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX Disclosure; SEQ ID XX; 284pp; Japanese.
XX The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences
SQ Sequence 602 BP; 148 A; 145 C; 128 G; 181 T; 0 U; 0 Other;
Query Match 79.0%; Score 15.8; DB 9; Length 602;
```



Best Local Similarity 89.5%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGGTGGGCAATCGT 19  
|||||  
DB 530 TCGAAGGGGAGCAATCGT 512

## RESULT 9

ACLI9645/C  
ID ACLI9645 standard; DNA; 622 BP.

XX ACCLI9645;

DT 17-OCT-2003 (first entry)

DE DNA clone originating in barley containing SNP encoding sequence #9636.

XX Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
KW gene; ss.

XX Hordeum vulgare.

XX WO2003057877-A1.

PD 17-JUL-2003.

PF 16-DEC-2002; 2002WO-1B005403.

PR 20-DEC-2001; 2001JP-00387059.

PR 20-DEC-2001; 2001JP-00387131.

PR 20-DEC-2001; 2001JP-00403299.

PR 20-DEC-2001; 2001JP-00403300.

PR 27-SEP-2002; 2002JP-00327515.

XX (UYN1-) UNITV JAPAN OKAYAMA.

XX Sato K, Takeda K, Kohara Y;

XX WPI; 2003-587127/55.

PT Single nucleotide polymorphism sites in barley varieties and DNA  
PT sequences containing them for analysis and identification of barley  
PT varieties and production of barley transformants with desired  
PT characteristics.

PS Disclosure; SEQ ID XX; 284bp; Japanese.

XX The present invention relates to oligonucleotide clones originating in  
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
CC varieties, identification of particular varieties and genotype-phenotype  
CC analysis, isolation of specific genes and creation of new varieties by  
CC transformation of barley varieties with them and production of new barley  
CC varieties with desired properties. The present sequence represents an  
CC oligonucleotide clone DNA sequence featured in the specification. The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published-pct-sequences

XX Sequence 622 BP; 151 A; 154 C; 132 G; 185 T; 0 U; 0 Other;

QY Query Match 79.0%; Score 15.8; DB 9; Length 622;

XX Best Local Similarity 89.5%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGGTGGGCAATCGT 19  
|||||

DB 540 TCGAAGGGGAGCAATCGT 522

RESULT 10  
ABSS53602/C

ID ABSS53602 standard; cDNA; 663 BP.

XX ABSS53602;

DT 29-NOV-2002 (first entry)

DE Human cDNA encoding Tyrosine specific protein phosphatase #2.

XX Human; ss; gene; tyrosine specific protein phosphatase; PTPase;  
KW central nervous system disorder; Parkinson's disease; gout; diabetes;  
KW Alzheimer's disease; dementia; motor neuron disease; multiple sclerosis;  
KW Pick's disease; Creutzfeldt-Jacob dementia; schizophrenia; pain;  
KW cognitive disorder; obesity; cachexia; anorexia; wasting disorders;  
KW hypertension; stroke; osteoarthritis; breast cancer; reduced fertility;  
KW pregnancy complication; depression.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS

XX /tag= a  
XX /product= "Tyrosine specific protein phosphatase"  
XX /EC\_number= "3.1.3.48"

XX WO200263014-A2.

PD 15-AUG-2002.

PF 04-FEB-2002; 2002WO-EP001099.

PR 05-FEB-2001; 2001US-0265872P.

PR (FARB ) BAYER AG.

XX Koehler RH;

XX WPI; 2002-627563/67.

XX P-PEDB; ABG32808.

PT New isolated polynucleotide encoding a human tyrosine-specific protein  
PT phosphatase (PTPase), useful for treating diseases associated with PTPase  
PT dysfunction e.g. obesity or diabetes.

PS Claim 1; Fig 1; 109pp; English.

XX The invention relates to an isolated polynucleotide encoding a tyrosine -  
CC specific protein phosphatase polypeptide (PTPase, appearing as ABSS53601  
CC and ABSS53602), a polynucleotide which hybridizes under stringent  
CC conditions to the PTPase nucleic acid, or a polynucleotide having a  
CC sequence that deviates from PTPase nucleic acid due to the degeneration  
CC of the genetic code. Also included are PTPase expression vectors,  
CC transformed host cells, the purified PTPase, producing the PTPase,  
CC detecting the PTPase or its nucleic acid, and screening for agents that  
CC reduce or modulate the activity of PTPase. The expression vector or the  
CC agent is useful for treating diseases associated with tyrosine-specific  
CC protein phosphatase dysfunction e.g. a central nervous system disorder  
CC (e.g. Parkinson's disease, Alzheimer's disease, dementia, motor neuron  
CC disease, multiple sclerosis, Pick's disease, Creutzfeldt-Jacob dementia,  
CC schizophrenia, pain and cognitive disorders), obesity cachexia, anorexia,  
CC wasting disorders, diabetes, hypertension, stroke, gout, osteoarthritis,  
CC breast cancer, reduced fertility, pregnancy complications and depression.  
CC Many other diseases and disorders are listed in the specification. The  
CC present sequence encodes a human PTPase

XX Sequence 663 BP; 146 A; 199 C; 214 G; 104 T; 0 U; 0 Other;

QY Query Match 79.0%; Score 15.8; DB 6; Length 663;

XX Best Local Similarity 89.5%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGGTGGGCAATCGT 19  
|||||

DB 359 TCGAAGGTGGGAGGTCGT 341

```
RESULT 11
ACLI19647/c
ID ACLI19647 standard; DNA; 714 BP.
AC
ACLI19647;
XX
XX 17-OCT-2003 (first entry)
XX
XX DNA clone originating in barley containing SNP encoding sequence #9638.
XX
XX Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
XX gene; ss.
XX
XX Hordeum vulgare.
XX
XX MO2003057877-A1.
XX
XX 17-JUL-2003.
XX
XX 16-DEC-2002; 2002MO-IB005403.
XX
XX 20-DEC-2001; 2001JP-00387059.
XX
XX 20-DEC-2001; 2001JP-00387131.
XX
XX 20-DEC-2001; 2001JP-00403299.
XX
XX 20-DEC-2001; 2001JP-00403300.
XX
XX 27-SEP-2002; 2002JP-00327515.
XX
XX (UYN1-) UNIV JAPAN OKAYAMA.
XX
XX Sato K, Takeda K, Kohara Y;
XX
XX WPI; 2003-587127/55.
XX
XX Single nucleotide polymorphism sites in barley varieties and DNA
XX sequences containing them for analysis and identification of barley
XX varieties and production of barley transformants with desired
XX characteristics.
XX
XX Disclousure: SEQ ID XX; 284pp; Japanese.
XX
XX The present invention relates to oligonucleotide clones originating in
XX barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX varieties, identification of particular varieties and genotype-phenotype
XX analysis, isolation of specific genes and creation of new varieties by
XX transformation of barley varieties with them and production of new barley
XX varieties with desired properties. The present sequence represents an
XX oligonucleotide clone DNA sequence featured in the specification. The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published-pct-sequences
XX
XX Sequence 714 BP; 190 A; 165 C; 150 G; 207 T; 0 U; 2 Other;
XX
XX Query Match 79.0%; Score 15.8; DB 9; Length 714;
XX Best Local Similarity 89.5%; Pred. No. 3.6e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 TCGAAGGTGGCAATCGT 19
XX |||||
XX 688 TCGAAGGGGAGCAATCGT 670
XX
XX
XX RESULT 12
XX ABS53601/c
XX ID ABS53601 standard; CDNA; 837 BP.
XX
XX ABS53601;
XX
XX 29-NOV-2002 (first entry)
XX
```

```
DE Human cDNA encoding Tyrosine specific protein phosphatase #1.
XX
XX Human; ss; gene; tyrosine specific protein phosphatase; PTPase;
XX central nervous system disorder; Parkinson's disease; gout; diabetes;
XX Alzheimer's disease; dementia; motor neuron disease; multiple sclerosis;
XX Pick's disease; Creutzfeldt-Jacob dementia; schizophrenia; pain;
XX cognitive disorder; obesity; cachexia; anorexia; wasting disorder;
XX hypertension; stroke; osteoarthritis; breast cancer; reduced fertility;
XX pregnancy complication; depression.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 46..837
XX FT /tag=a
XX FT /product="Tyrosine specific protein phosphatase"
XX FT /EC_number="3.1.3.48"
XX
XX MO200263014-A2.
XX
XX 15-AUG-2002.
XX
XX 04-FEB-2002; 2002MO-EP001099.
XX
XX 05-FEB-2001; 2001US-0265872P.
XX
XX (FARB ) BAYER AG.
XX
XX Koehler RH;
XX
XX WPI; 2002-627563/67.
XX
XX P-PSDB; ABG32807.
XX
XX New isolated polynucleotide encoding a human tyrosine-specific protein
XX phosphatase (PTPase), useful for treating diseases associated with PTPase
XX dysfunction e.g. obesity or diabetes.
XX
XX Claim 1; Fig 5; 109pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding a tyrosine
XX specific protein phosphatase polypeptide (PTPase, appearing as ABS53601
XX and ABS53602), a polynucleotide which hybridizes under stringent
XX conditions to the PTPase nucleic acid, or a polynucleotide having a
XX sequence that deviates from PTPase nucleic acid due to the degeneration
XX of the genetic code. Also included are PTPase expression vectors,
XX transformed host cells, the purified PTPase, producing the PTPase,
XX detecting the PTPase or its nucleic acid, and screening for agents that
XX reduce or modulate the activity of PTPase. The expression vector or the
XX agent is useful for treating diseases associated with tyrosine-specific
XX protein phosphatase dysfunction e.g. a central nervous system disorder
XX (e.g. Parkinson's disease, Alzheimer's disease, dementia, motor neuron
XX disease, multiple sclerosis, Pick's disease, Creutzfeldt-Jacob dementia,
XX schizophrenia, pain and cognitive disorders), obesity cachexia, anorexia,
XX wasting disorders, diabetes, hypertension, stroke, gout, osteoarthritis,
XX breast cancer, reduced fertility, pregnancy complications and depression.
XX Many other diseases and disorders are listed in the specification. The
XX present sequence encodes a human PTPase
XX
XX Sequence 837 BP; 203 A; 231 C; 252 G; 151 T; 0 U; 0 Other;
XX
XX Query Match 79.0%; Score 15.8; DB 6; Length 837;
XX Best Local Similarity 89.5%; Pred. No. 3.7e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 TCGAAGGTGGCAATCGT 19
XX |||||
XX 533 TCGAAGGTGGCAATCGT 515
XX
XX
XX RESULT 13
XX ABZ12053/c
XX ID ABZ12053 standard; CDNA; 950 BP.
XX
```

AC AB212053;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 935.  
XX  
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
cell-proliferative disorder; neurodegenerative disease; bacterial;  
Parkinson's disease; Alzheimer's disease; autoimmune disease;  
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
arthritis; cytostatic; immunomodulator; nocitropic; neuroprotective;  
antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
antiarthritic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200270539-A2.  
XX  
PD 12-SEP-2002.  
XX  
PP 05-MAR-2002; 2002WO-US005095.  
XX  
PR 05-MAR-2001; 2001US-00799451.  
XX  
PA (HYSE-) HYSEQ INC.  
PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao Q, Ren F;  
PI Xue AU, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M,  
PI Wehman T, Wang J, Wang D, Drmanac RT;  
DR WPI; 2002-759812/82.  
DR P-PsDB; ABP69836.  
XX  
PT New polynucleotides comprising sequences assembled from expressed  
sequence tags (ESTs), useful for treating cell-proliferative,  
neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
or coagulation disorders.  
XX  
PS Claim 1; SEQ ID NO 935; 1012pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated polynucleotide (1) comprising a  
nucleotide sequence selected from any of 948 sequences (AB21119-  
AB212066) or their mature protein coding portion, active domain coding  
protein or complementary sequences. The polynucleotides are useful for  
identifying expressed genes or for physical mapping of human genome. The  
encoded polypeptides (ABP6902-ABP69849) are useful as molecular weight  
markers, as a food supplement, for generating antibodies, in medical  
imaging, screening and diagnostic assays and for treating cell-  
proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
platelet or coagulation disorders, wound, burn, incision, ulcers, liver  
or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
arthritis, etc. Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic format  
directly from WIGO at ftp.wigo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 950 BP; 212 A; 279 C; 302 G; 157 T; 0 U; 0 Other;  
XX  
Query Match 79.0%; Score 15.8; DB 6; Length 950;  
Best Local Similarity 89.5%; Pred. No. 3.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TCGAAGTGGCGCAATCGT 19  
| | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | |  
Db 645 TCGAAGTGGCGCAGTCTG 627  
| | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | |  
RESULT 14  
ADM44571/c  
ID ADM44571 standard; cDNA; 950 BP.  
XX

AC ADM44571;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Novel human arginine-rich protein cDNA #935.  
XX  
KW ss; gene; human; arginine-rich protein; cancer; inflammation;  
genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN US2004053250-A1.  
XX  
PD 18-MAR-2004.  
XX  
PP 21-NOV-2002; 2002US-00302172.  
XX  
PR 05-MAR-2001; 2001US-00799451.  
PR 05-MAR-2002; 2002WO-US005095.  
PR 20-AUG-2002; 2002US-00225251.  
XX  
PA (TANG/) TANG Y T.  
PA (XUE/) XUE A.  
PA (DRMA/) DRMANAC R T.  
PI Tang YT, Xue A, Drmanac RT;  
PI WPI; 2004-238579/22.  
DR  
PT New isolated arginine-rich protein-like polynucleotides and polypeptides,  
useful for diagnosing and/or treating conditions associated with aberrant  
PT activity of the arginine-rich polypeptides, such as cancer and  
PT inflammation.  
XX  
PS Disclosure; SEQ ID NO 935; 51pp; English.  
XX  
CC The invention relates to an isolated polynucleotide. The methods and  
compositions of the present invention are useful for the diagnosis and/or  
treatment of diseases or conditions associated with aberrant expression  
or activity of the arginine-rich protein-like polypeptides, such as  
cancer and inflammation. They can also be used in forensics, gene  
CC mapping, identification of mutations responsible for genetic disorders,  
CC and in assessing biodiversity. The present sequence represents a novel  
human arginine-rich protein cDNA.  
XX  
SQ Sequence 950 BP; 212 A; 279 C; 302 G; 157 T; 0 U; 0 Other;  
XX  
Query Match 79.0%; Score 15.8; DB 12; Length 950;  
Best Local Similarity 89.5%; Pred. No. 3.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TCGAAGTGGCGCAATCGT 19  
| | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | |  
Db 645 TCGAAGTGGCGCAGTCTG 627  
| | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | |  
RESULT 15  
ADFS3637/c  
ID ADFS3637 standard; DNA; 1104 BP.  
XX  
AC ADFS3637;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Murine DNA encoding phosphatase protein, SEQ ID NO 5.  
XX  
KW phosphatase; mutated; computer readable recording medium; solid support;  
ds; mouse; murine; gene.  
XX  
OS Mus musculus.  
XX  
PN WO2003091428-A1.  
XX

```

PD 06-NOV-2003.
XX
XX 23-APR-2003; 2003MO-JP005175.
XX
XX 23-APR-2002; 2002JP-00120709.
PR 04-DEC-2002; 2002JP-00352308.
XX
XX (RIKE ) RIKEN KK.
XX (DNAF-) DNAFORM KK.
PA (MITU ) MITSUBISHI CHEM CORP.
XX
XX Hayashizaki Y, Kamiya M, Kubodera H;
XX WPI; 2003-854398/79.
DR P-PSDB; ADF53644.
XX
XX Proteins with phosphatase activity useful for screening for substances
XX that affect their activity.
XX
XX Claim 4; SEQ ID NO 5; 170pp; Japanese.
XX
XX The invention relates to novel isolated proteins comprising an amino acid
XX sequence represented by any of SEQ ID Nos: 8 to 14, 25 and 26; and a
XX protein comprising an amino acid sequence derived from an amino acid
XX sequence represented by any of SEQ ID Nos: 8 to 14, 25 and 26 by
XX deletion, substitution and/or addition of one to several amino acids and
XX having a phosphatase activity. The invention further comprises: the DNA
XX of optionally mutated sequences SEQ ID Nos: 1-7, 23 and 24 or those that
XX hybridise to them; DNA and cDNA encoding the proteins; vectors containing
XX the DNA; cells containing the vectors; proteins produced by the cells;
XX whole or fragments of antibodies against the proteins; screening for
XX substances that affect the expression and activity of the proteins; a
XX computer readable recording medium; and a solid support bound to the DNA.
XX The novel phosphatase proteins are useful for screening for substances
XX that affect the activity of the proteins, which have potential use in the
XX development of drugs for treating associated disorders. This
XX polynucleotide represents the DNA of an optionally mutated sequence of
XX the invention.
XX
SQ Sequence 1104 BP; 255 A; 267 C; 333 G; 229 T; 0 U; 0 Other;

```

```

Query Match          79.0%; Score 15.8; DB 10; Length 1104;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 TCGAAGTGGGCAATCGT 19
      |||||
Db      453 TCGAAGTGGGCAATCGT 435

```

Search completed: November 7, 2005, 05:40:19  
 Job time : 237.638 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 19:19:37 / Search time 67.8723 Seconds  
(without alignments)  
482.163 Million cell updates/sec

Title: US-10-777-131A-3

Perfect score: 20

Sequence: 1 tcgaagtggtgccaatcgtg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405566

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: Issued Patents NA:

1: /cgn2\_6/ptodata/1/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6C COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/6D COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
2	20	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
3	15.8	79.0	601	4 US-09-949-016-117629	Sequence 117629,
4	15.8	79.0	601	4 US-09-949-016-117655	Sequence 117655,
5	15.8	79.0	601	4 US-09-949-016-141204	Sequence 141204,
6	15.8	79.0	950	4 US-09-799-451-935	Sequence 935, App
7	15.8	79.0	1422	4 US-09-701-896-1	Sequence 971, App
8	15.8	79.0	3615	3 US-09-221-017B-971	Sequence 971, App
9	15.8	79.0	16112	4 US-09-949-016-15671	Sequence 16671, A
10	15.8	79.0	18264	4 US-09-949-016-15469	Sequence 16469, A
11	15.8	79.0	33753	4 US-09-949-016-15741	Sequence 15741, A
12	15.8	79.0	33756	4 US-09-949-016-12006	Sequence 12006, A
13	15.8	79.0	46288	4 US-09-949-016-14999	Sequence 14999, A
14	15.8	79.0	46311	4 US-09-949-016-15000	Sequence 15000, A
15	15.8	79.0	113876	4 US-09-949-016-14828	Sequence 14828, A
16	15.8	79.0	113876	4 US-09-949-016-14829	Sequence 14829, A
17	15.8	79.0	115508	4 US-09-949-016-11800	Sequence 11800, A
18	15.8	79.0	115508	4 US-09-949-016-14826	Sequence 14826, A
19	15.8	79.0	115508	4 US-09-949-016-14827	Sequence 14827, A
20	15.4	77.0	12042	4 US-09-949-016-14331	Sequence 13431, A
21	15.4	77.0	12797	4 US-09-949-016-15274	Sequence 15274, A
22	15.4	77.0	15499	4 US-09-949-016-14690	Sequence 14690, A
23	15.4	77.0	15499	4 US-09-949-016-14691	Sequence 14691, A
24	15.4	77.0	15499	4 US-09-949-016-14692	Sequence 14692, A
25	15.4	77.0	44789	4 US-09-949-016-13909	Sequence 13909, A
26	15.4	77.0	45693	4 US-09-949-016-12523	Sequence 12523, A
27	15.4	77.0	45693	4 US-09-949-016-12678	Sequence 12678, A

C 28	15.4	77.0	45693	4	US-09-949-016-12679	Sequence 12679, A
C 29	15.4	77.0	50266	4	US-09-949-016-17598	Sequence 17598, A
C 30	15.2	76.0	365	4	US-09-270-767-28333	Sequence 28333, A
C 31	15.2	76.0	773	4	US-09-270-767-13748	Sequence 13748, A
C 32	15.2	76.0	780	4	US-09-724-623-156	Sequence 56, Appl
C 33	15.2	76.0	846	4	US-09-583-110-2248	Sequence 2248, Ap
C 34	15.2	76.0	846	4	US-09-107-433-470	Sequence 470, App
C 35	15.2	76.0	934	4	US-09-270-767-12547	Sequence 12547, A
C 36	15.2	76.0	1542	3	US-09-385-028-16	Sequence 16, Appl
C 37	15.2	76.0	1542	4	US-09-726-614-16	Sequence 16, Appl
C 38	15.2	76.0	1542	4	US-09-385-040-16	Sequence 16, Appl
C 39	15.2	76.0	11604	3	US-09-385-028-13	Sequence 13, Appl
C 40	15.2	76.0	11604	4	US-09-726-614-13	Sequence 13, Appl
C 41	15.2	76.0	11604	4	US-09-385-028-13	Sequence 13, Appl
C 42	15.2	76.0	15079	3	US-09-385-028-1	Sequence 1, Appl
C 43	15.2	76.0	15079	4	US-09-726-614-1	Sequence 1, Appl
C 44	15.2	76.0	15120	4	US-09-385-040-1	Sequence 1, Appl
C 45	15.2	76.0	36412	4	US-08-311-731A-132	Sequence 132, App

#### ALIGNMENTS

```
RESULT 1
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
Query Match          100.0%; Score 20; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. NO. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 4385490 TCGAAGTGGCGCAATCTGTG 4385509
OY 1 TCGAAGTGGCGCAATCTGTG 20
|||||
DB 4385490 TCGAAGTGGCGCAATCTGTG 4385509
RESULT 2
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
```

```
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match          100.0%; Score 20; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1  TCGAAGTGGCGCAATCGTG 20
Db      4393252 TCGAAGTGGCGCAATCGTG 4393271
```

```
RESULT 3
US-09-949-016-117629/c
; Sequence 117629, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117629
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-117629
```

```
Query Match          79.0%; Score 15.8; DB 4; Length 601;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      2  CGAAGTGGCGCAATCGTG 20
Db      322  CGAAGTGGCGCAATCATG 304
```

```
RESULT 4
US-09-949-016-117655/c
; Sequence 117655, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117655
; LENGTH: 601
; TYPE: DNA
```

```
; ORGANISM: Human
US-09-949-016-117655
```

```
Query Match          79.0%; Score 15.8; DB 4; Length 601;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      2  CGAAGTGGCGCAATCGTG 20
Db      322  CGAAGTGGCGCAATCATG 304
```

```
RESULT 5
US-09-949-016-141204
; Sequence 141204, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 141204
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-141204
```

```
Query Match          79.0%; Score 15.8; DB 4; Length 601;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      2  CGAAGTGGCGCAATCGTG 20
Db      239  CGAAGTGGCGCAGATCGTG 257
```

```
RESULT 6
US-09-799-451-935/c
; Sequence 935, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyun
; APPLICANT: Zhang, Jie
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ma, Yundong
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Yang, Yonghong
; APPLICANT: Wenman, Tom
; APPLICANT: Ghosh, Reena
; APPLICANT: Drmanac, Radote T.
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803
; CURRENT APPLICATION NUMBER: US/09/799,451
```

CURRENT FILING DATE: 2001-03-05  
 NUMBER OF SEQ ID NOS: 948  
 SOFTWARE: pf\_fl\_genes Version 2.0  
 SEQ ID NO 935  
 LENGTH: 950  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (287)..(946)  
 US-09-799-451-935

Query Match 79.0%; Score 15.8; DB 4; Length 950;  
 Best Local Similarity 89.5%; Pred. No. 1e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGT 19  
 Db 645 TCGAAGTGGGCAATCGT 627

RESULT 7  
 US-09-701-896-1  
 Sequence 1, Application US/09701896  
 Patent No. 6706269  
 GENERAL INFORMATION:  
 APPLICANT: Jean-Louis Ruelie  
 APPLICANT: Carlot Vinals y de Bascois  
 TITLE OF INVENTION: No. 6706269el Compounds  
 FILE REFERENCE: BM45325  
 CURRENT APPLICATION NUMBER: US/09/701,896  
 CURRENT FILING DATE: 2002-02-26  
 PRIOR APPLICATION NUMBER: 9812163.5  
 PRIOR FILING DATE: 1998-06-05  
 NUMBER OF SEQ ID NOS: 14  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1  
 LENGTH: 1422  
 TYPE: DNA  
 ORGANISM: Moraxella catarrhalis  
 US-09-701-896-1

Query Match 79.0%; Score 15.8; DB 4; Length 1422;  
 Best Local Similarity 89.5%; Pred. No. 1e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGT 19  
 Db 1244 TCGAAGTGGGCAATCGT 1262

RESULT 8  
 US-09-221-017B-971/c  
 Sequence 971, Application US/09221017B  
 Patent No. 6444799  
 GENERAL INFORMATION:  
 APPLICANT: ROSE, Bruce C.  
 TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
 NUMBER OF SEQUENCES: 1120  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MORRISON & FOERSTER  
 STREET: 755 PAGE MILL ROAD  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304-1018  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows  
 SOFTWARE: FastSeq for Windows Version 2.0b  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PP1182  
 FILING DATE: 31-DEC-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PP1546  
 FILING DATE: 30-JAN-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PP2911  
 FILING DATE: 09-APR-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/AU98/01023  
 FILING DATE: 10-DEC-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Montoy, Gladys H  
 REGISTRATION NUMBER: 32,430  
 REFERENCE/DOCKET NUMBER: 27340-20021.00  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-813-5600  
 TELEFAX: 650-494-0792  
 TELEX: 706141  
 INFORMATION FOR SEQ ID NO: 971:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3615 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: circular  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: UNKNOWN  
 ORIGINAL SOURCE:  
 ORGANISM: PORPHYROMONAS GINGIVALIS  
 FEATURES:  
 NAME/KEY: misc\_feature  
 LOCATION: 1...3615  
 US-09-221-017B-971

Query Match 79.0%; Score 15.8; DB 3; Length 3615;  
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGT 19  
 Db 421 TCGAATGGGCAATCT 403

RESULT 9  
 US-09-949-016-16671/c  
 Sequence 16671, Application US/09949016  
 Patent No. 6812339  
 GENERAL INFORMATION:  
 APPLICANT: VENTER, J. Craig et al.  
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 FILE REFERENCE: C1001307  
 CURRENT APPLICATION NUMBER: US/09/949,016  
 CURRENT FILING DATE: 2000-04-14  
 PRIOR APPLICATION NUMBER: 60/241,755  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/237,768  
 PRIOR FILING DATE: 2000-10-03  
 PRIOR APPLICATION NUMBER: 60/231,498  
 PRIOR FILING DATE: 2000-09-08  
 NUMBER OF SEQ ID NOS: 207012  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 16671  
 LENGTH: 16112  
 TYPE: DNA  
 ORGANISM: Human  
 US-09-949-016-16671  
 Query Match 79.0%; Score 15.8; DB 4; Length 16112;





```
; Sequence 15000, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 15000
; LENGTH: 46311
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15000
```

```
Query Match 79.0%; Score 15.8; DB 4; Length 46311;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 2 CGAAGTGGCAATCGTG 20
Db 10810 CCAAGTGGCAATCATG 10828
```

```
RESULT 15
US-09-949-016-14828
; Sequence 14828, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 14828
; LENGTH: 113876
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14828
```

```
Query Match 79.0%; Score 15.8; DB 4; Length 113876;
Best Local Similarity 89.5%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 2 CGAAGTGGCAATCGTG 20
Db 79450 CCAAGTGGCAATCATG 79468
```

```
Search completed: November 6, 2005, 23:10:27
Job time : 85.8723 secs
```

This page blank (spec)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 20:56:42 ; Search time 498.723 Seconds  
(without alignments)  
331.640 Million cell updates/sec

Title: US-10-777-131A-3

Perfect score: 20

Sequence: 1 tcgaagctggcgcaatcgtg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:\*  
23: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
24: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/2/pubpna/US11\_PUBCOMB.seq:\*  
26: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
27: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
28: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16	US-10-216-817-3
2	20	100.0	20	22	US-10-777-131A-3
3	20	100.0	1144	16	US-10-216-817-20
4	20	100.0	1144	22	US-10-777-131A-20
5	17.4	87.0	937	20	US-10-437-963-37441

Query Match	Score	DB	Length	Sequence
100.0%	20	16	Sequence 3, Appl1	
100.0%	20	22	Sequence 3, Appl1	
100.0%	1144	16	Sequence 20, Appl1	
100.0%	1144	22	Sequence 20, Appl1	
87.0%	937	20	Sequence 37441, A	

#### ALIGNMENTS

RESULT 1  
US-10-216-817-3  
; Sequence 3, Application US/10216817  
; Publication No. US20030129619A1  
GENERAL INFORMATION:  
APPLICANT: GICQUEL, BRIGITTE  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
FILE REFERENCE: 03495.0233-00000  
CURRENT APPLICATION NUMBER: US/10/216,817  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 60/311,824  
PRIOR FILING DATE: 2001-08-14  
PRIOR APPLICATION NUMBER: 60/313,523  
PRIOR FILING DATE: 2001-08-21  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 20  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Primer  
US-10-216-817-3  
Query Match 100.0%; Score 20; DB 16; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.7;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGGTGGGCAAAATCGTG 20  
Db 1 TCGAAGGTGGGCAAAATCGTG 20

## RESULT 2

US-10-777-131A-3  
; Sequence 3, Application US/10777131A  
; Publication No. US20050026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; TITLE OF INVENTION: IN GENES OF THE MUTC FAMILY  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 3  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer  
US-10-777-131A-3

Query Match 100.0%; Score 20; DB 22; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGGTGGGCAAAATCGTG 20  
Db 1 TCGAAGGTGGGCAAAATCGTG 20

## RESULT 3

US-10-216-817-20  
; Sequence 20, Application US/10216817  
; Publication No. US20030129619A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; TITLE OF INVENTION: IN GENES OF THE MUTC FAMILY  
; FILE REFERENCE: 03495.0233-00000  
; CURRENT FILING DATE: US/10/216,817  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 1144  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-216-817-20

Query Match 100.0%; Score 20; DB 16; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGGTGGGCAAAATCGTG 20

Db 7 TCGAAGGTGGGCAAAATCGTG 26

## RESULT 4

US-10-777-131A-20  
; Sequence 20, Application US/10777131A  
; Publication No. US20050026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; TITLE OF INVENTION: IN GENES OF THE MUTC FAMILY  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT FILING DATE: US/10/777,131A  
; CURRENT FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 20  
; LENGTH: 1144  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-777-131A-20

Query Match 100.0%; Score 20; DB 22; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGGTGGGCAAAATCGTG 20  
Db 7 TCGAAGGTGGGCAAAATCGTG 26

## RESULT 5

US-10-437-963-37441/C  
; Sequence 37441, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Zhou, Yinhua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbasuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT FILING DATE: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 37441  
; LENGTH: 937  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_41171C.1  
US-10-437-963-37441

Query Match 87.0%; Score 17.4; DB 20; Length 937;  
Best Local Similarity 94.7%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGAAGGTGGGCAAAATCGT 19  
Db 758 TCGAAGGTGGGCAAAATCAT 740

```
RESULT 6
US-10-425-115-44578/c
; Sequence 44578, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 44578
; LENGTH: 241
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_14065C.1
US-10-425-115-44578

Query Match      84.0%; Score 16.8; DB 21; Length 241;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TCGAAGTGGGCAATCGTG 20
      |||||
DB      99 TTGAAGGTGGGAATTCGTG 80

RESULT 7
US-10-425-115-175990
; Sequence 175990, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 175990
; LENGTH: 2589
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) (2589)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_92098C.1
US-10-425-115-175990

Query Match      84.0%; Score 16.8; DB 21; Length 2589;
Best Local Similarity 90.0%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TCGAAGTGGGCAATCGTG 20
      |||||
DB      1372 TCGAAGGTGTCATATCGTG 1391

RESULT 8
US-10-425-115-175994
; Sequence 175994, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 175994
; LENGTH: 5658
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_92100C.1
US-10-425-115-175994

Query Match      84.0%; Score 16.8; DB 21; Length 5658;
Best Local Similarity 90.0%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TCGAAGTGGGCAATCGTG 20
      |||||
DB      1558 TCGAAGGTGTCATATCGTG 1577

RESULT 9
US-09-925-065A-830421
; Sequence 830421, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; CURRENT APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 830421
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Clone ID: MRT4577_92100C.1
US-09-925-065A-830421

Query Match      82.0%; Score 16.4; DB 13; Length 588;
Best Local Similarity 94.4%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GAAAGTGGGCAATCGTG 20
      |||||
DB      144 GAAAGTGGGCAATCGTG 161

RESULT 10
US-10-424-599-105971/c
; Sequence 105971, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

```

; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongshui
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 105971
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_66708C.1
US-10-424-599-105971

Query Match      82.0%; Score 16.4; DB 19; Length 1065;
Best Local Similarity 94.4%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy      3  CGAAGTGGGCAATCGTG 20
Db      857  GAAAGTGGGCAATAGTG 840

RESULT 11
US-10-220-891-20
; Sequence 20, Application US/10220891
; Publication No. US20030207286A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWARA, AKIRA
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES HAVING CHARACTERISTICS OF ENHANCED
; TITLE OF INVENTION: EXPRESSION IN HUMAN NEUROBLASTOMA WITH FAVORABLE PROGNOSIS
; TITLE OF INVENTION: BASED ON COMPARISON BETWEEN HUMAN NEUROBLASTOMA WITH FAVORABLE
; TITLE OF INVENTION: PROGNOSIS AND HUMAN NEUROBLASTOMA WITH UNFAVORABLE PROGNOSIS
; FILE REFERENCE: 7388-73435
; CURRENT APPLICATION NUMBER: US/10/220,891
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: JP 2000/140387
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: JP 2000/159195
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 2267
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-220-891-20

Query Match      80.0%; Score 16; DB 18; Length 2267;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      2  CGAAGTGGGCAATC 17
Db      1970  CGAAGTGGGCAATC 1985

RESULT 12
US-09-925-065A-140048
; Sequence 140048, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
```

```

; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140048
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-140048

Query Match      79.0%; Score 15.8; DB 13; Length 484;
Best Local Similarity 89.5%; Pred. No. 7.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      2  CGAAGTGGGCAATCGTG 20
Db      123  CGAAGTGGGCAATCATG 141

RESULT 13
US-10-972-079-52425/c
; Sequence 52425, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF
; TITLE OF INVENTION: LIVESTOCK
; FILE REFERENCE: MM11110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52425
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894303000_4
US-10-972-079-52425

Query Match      79.0%; Score 15.8; DB 24; Length 600;
Best Local Similarity 89.5%; Pred. No. 7.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      2  CGAAGTGGGCAATCGTG 20
Db      500  CGAAGTGGGCAAAACCTG 482

RESULT 14
US-11-060-756-3208/c
; Sequence 3208, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
```

```

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3208
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-3208

```

```

Query Match          79.0%; Score 15.8; DB 26; Length 600;
Best Local Similarity 89.5%; Pred. No. 7.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 TCGAAGGTGGGCAATCGT 19
        |||||
Db      161 TCGAAGGTGGGCAATCGT 143

```

```

RESULT 15
US-11-060-756-7480/c
; Sequence 7480, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7480
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-7480

```

```

Query Match          79.0%; Score 15.8; DB 26; Length 600;
Best Local Similarity 89.5%; Pred. No. 7.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 TCGAAGGTGGGCAATCGT 19
        |||||
Db      161 TCGAAGGTGGGCAATCGT 143

```

Search completed: November 7, 2005, 05:21:11  
Job time : 501.723 secs

THE DOG PARK (1994)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 2000.85 Seconds  
(without alignments)  
380.481 Million cell updates/sec

Title: US-10-777-131A-3

Perfect score: 20

Sequence: 1 tcgaagtggtggaatcgtg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.4	92.0	592	7	CV251860 MS0118.B2
C 2	17.4	87.0	633	8	AZ564202 RPT-23-2
C 3	17	85.0	559	7	CK124820 BES182411
C 4	17	85.0	790	2	BF627004 HVSMB000
C 5	16.8	84.0	391	9	CG049330 PUF044TD
C 6	16.8	84.0	452	9	CL219748 ZMMB044
C 7	16.8	84.0	511	9	CC803612 ih17b02.9
C 8	16.8	84.0	548	4	CC803612 ih17b02.9
C 9	16.8	84.0	553	5	BQ097494 ph03d06.y
C 10	16.8	84.0	607	8	AZ471085 IM0285D02
C 11	16.8	84.0	608	4	BJ460007 BJ460007
C 12	16.8	84.0	627	1	AV910592 AV910592
C 13	16.8	84.0	642	4	BJ462359 BJ462359
C 14	16.8	84.0	646	4	BJ452459 BJ452459
C 15	16.8	84.0	658	1	AV920519 AV920519
C 16	16.8	84.0	664	8	AZ574524 329PFC09
C 17	16.8	84.0	678	4	BJ460029 BJ460029
C 18	16.8	84.0	681	7	CV232798 WS0198.B2
C 19	16.8	84.0	685	1	AV919237 AV919237
C 20	16.8	84.0	701	1	AV914195 AV914195
C 21	16.8	84.0	702	4	BJ448654 BJ448654
C 22	16.8	84.0	711	9	CC671127 OGUP72TV
C 23	16.8	84.0	728	4	BJ462046 BJ462046
C 24	16.8	84.0	803	8	BH574525 BOCJM13TF

25	16.8	84.0	827	9	CV538423	CV538423 OP_Ba001
26	16.8	84.0	828	9	CG004578	CG004578 ZUAAS80TH
27	16.8	84.0	843	8	BZ701269	BZ701269 PUCR48TD
C 28	16.8	84.0	848	4	B1116452	B1116452 602868878
C 29	16.8	84.0	941	4	BF965968	BF965968 602277143
C 30	16.8	84.0	947	4	CC412703	CC412703 PUNHW18TB
C 31	16.8	84.0	956	8	CC412704	CC412704 PUNHW18TD
C 32	16.8	84.0	1035	9	CL451811	CL451811 ZMMB047
C 33	16.8	84.0	1041	7	CF581680	CF581680 AGENCOURT
C 34	16.8	84.0	1172	8	CC268595	CC268595 CH261-67F
C 35	16.8	84.0	1604	2	BE613467	BE613467 601504233
C 36	16.4	82.0	328	1	AA861299	AA861299 ak33h05.x
C 37	16.4	82.0	329	1	AT425088	AT425088 t93h04.x
C 38	16.4	82.0	360	1	AT221296	AT221296 g974f01.x
C 39	16.4	82.0	417	5	B0831776	B0831776 T025F06.P
C 40	16.4	82.0	443	5	B0821612	B0821612 UB25CP08
C 41	16.4	82.0	484	9	CL876478	CL876478 abf13a1.1
C 42	16.4	82.0	490	1	A1161906	A1161906 A009P37U
C 43	16.4	82.0	498	7	CF230920	CF230920 PtAC0014H
C 44	16.4	82.0	511	1	AV699630	AV699630 AV699630
C 45	16.4	82.0	526	9	AG246983	AG246983 Lotus cor

#### ALIGNMENTS

RESULT 1  
LOCUS CV251860/c 592 bp mRNA linear EST 22-SEP-2004  
DEFINITION WS0118.B21\_F22 PT-P-FL-A-2 Populus balsamifera subsp. trichocarpa  
CDNA clone WS0118\_F22 3', mRNA sequence.

ACCESSION CV251860  
VERSION CV251860.1 GI:52504835  
KEYWORDS  
SOURCE  
ORGANISM

Populus balsamifera subsp. trichocarpa (Populus trichocarpa)  
Populus balsamifera subsp. trichocarpa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Malpighiales; Salicaceae; Salicaceae; Populus.

REFERENCE  
AUTHORS  
1 (bases 1 to 592)  
Raip, S., Cooper, D., Kolesova, N., Oddy, C., Butterfield, Y.,  
Kilpatrick, R., Liu, J., Palmquist, D., Stott, J., Barber, S., Yang, G.,  
Babakoff, R., Brown-John, M., Chand, S., Featherstone, R., Masson, A.,  
Mayo, M., Moran, J., Olson, T., Wong, D., Ritzland, C.E., Siddiqui, A.,  
Holt, R., Jones, S., Marra, M., Ellis, B.E., Douglas, C., Ritzland, K., and  
Bohlmann, J.

The poplar transcriptome: Analysis of expressed sequence tags from  
multiple cDNA libraries  
Unpublished (2004)

JOURNAL  
COMMENT  
Contact: Joerg Bohlmann  
Genome BC forest genomics program  
University of British Columbia  
UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237,  
Vancouver, British Columbia, Canada, V6T 1Z3  
Tel: 1-604-822-0282  
Fax: 1-604-822-0297

Email: bohlmann@interchange.ubc.ca  
Plate: WS0118 row: F column: 22  
High quality sequence stop: 592  
POLYA=yes.

FEATURES  
SOURCE  
Location/Qualifiers

1..592  
/organism="Populus balsamifera subsp. trichocarpa"  
/mol\_type="mRNA"  
/catalyze="383-2499 (Nisqually-1)"  
/sub\_species="trichocarpa"  
/db\_xref="taxon:3694"  
/clone="WS0118\_F22"  
/sex="Female"  
/lab\_host="E. coli DH10B T1 phage resistant cells"  
/clone\_id="PF-P-FL-A-2"  
/note="Vector: pBluescript II SK (+) XR; Site 1: SacI (5'  
end of cDNA); Site 2: XhoI (3' end of cDNA); Ploem and

cambium from 8 year old trees harvested within the Boise Cascade region of Washington state on May 15th, 2001. cDNA was prepared from 20 micrograms of mRNA according to the full-length cDNA library construction method described by Carninci P. et al. (2000), Genome Research 10(10):1617-1630 and directionally ligated into the Bluescript II SK (+) XR vector digested with SstI (5' end) and XhoI (3'). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation."

## ORIGIN

Query Match 92.0%; Score 18.4; DB 7; Length 592;  
Best Local Similarity 95.0%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGAAGTGGCGCAATCGT 20  
|||||  
545 TCGAAGTGGCGCAACTCGT 526

## RESULT 2

AZ564202 633 bp DNA linear GSS 20-NOV-2000  
LOCUS RPCI-23-208N19.TV RPCI-23 Mus musculus genomic clone  
DEFINITION RPCI-23-208N19, genomic survey sequence.  
ACCESSION AZ564202  
VERSION AZ564202.1 GI:11244022  
KEYWORDS GSS.

## SOURCE

Mus musculus (house mouse)  
Mus musculus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 633)

## REFERENCE

Zhao, S., Nierman, W., Feldbylum, T., Malek, J., Shateman, S., Akhter, B., Levin, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)

## AUTHORS

Other GSSs: RPCI-23-208N19.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

## TITLE

http://www.tigr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html  
Place: 208 row: N column: 19  
Seq primer: SP6  
Class: BAC ends.

## JOURNAL

## COMMENT

Email: szhae@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.bufileo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html  
Place: 208 row: N column: 19  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

## source

Location/Qualifiers  
1..633  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-208N19"  
/sex="Female"  
/lab\_host="DH10B"  
/clone\_11b="RPCI-23"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site\_1: SCOR1; Site\_2: SCOR1; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 633;  
Best Local Similarity 94.7%; Pred. No. 4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGAAGTGGCGCAATCGT 19  
|||||  
491 TCGAAGTGGCGCAATCGT 473

## RESULT 3

CK124820 559 bp mRNA linear EST 01-MAR-2004  
LOCUS BS1824110H21 BS1824 Hordeum vulgare subsp. vulgare cDNA clone  
DEFINITION MPMP2010H2110 5-PRIME, mRNA sequence.  
ACCESSION CK124820  
VERSION CK124820.1 GI:44807822  
KEYWORDS EST.

## SOURCE

Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.

## ORGANISM

1 (bases 1 to 559)  
Kramer, A., Fellner, T., Possing, A., Radchuk, V., Weschke, W., Buerkle, L. and Kersten, B.  
Application of the protein microarray technology for the identification of expression library derived target proteins for barley protein kinase CK2  
Unpublished (2003)

## REFERENCE

Contact: Birgit Kersten\* and Winfriede Weschke\*\*  
\*Plant Protein Chip Group, Department Lebrach, \*\*Department Molecular Genetics, Gene Expression Group  
\*Max-Planck-Institute for Molecular Genetics, \*\*Institute of Plant Genetics and Crop Plant Research Gatersleben  
\*Innestr. 73, D-14195 Berlin, Germany, \*\*Corrensstrasse 3, D-06466 Gatersleben, Germany  
Tel: \*\*49(0)30/84131648, \*\*49(0)394825500  
Fax: \*\*49(0)30/84131128, \*\*49(0)3948255237  
Email: \*kersten@molgen.mpg.de, \*\*weschke@ipk-gatersleben.de  
Insert Length: 559 Std Error: 0.00  
Plate: 10 row: H column: 21  
Seq primer: POB65.

## TITLE

Identification of the protein microarray technology for the identification of expression library derived target proteins for barley protein kinase CK2  
Unpublished (2003)

## JOURNAL

## COMMENT

Contact: Birgit Kersten\* and Winfriede Weschke\*\*  
\*Plant Protein Chip Group, Department Lebrach, \*\*Department Molecular Genetics, Gene Expression Group  
\*Max-Planck-Institute for Molecular Genetics, \*\*Institute of Plant Genetics and Crop Plant Research Gatersleben  
\*Innestr. 73, D-14195 Berlin, Germany, \*\*Corrensstrasse 3, D-06466 Gatersleben, Germany  
Tel: \*\*49(0)30/84131648, \*\*49(0)394825500  
Fax: \*\*49(0)30/84131128, \*\*49(0)3948255237  
Email: \*kersten@molgen.mpg.de, \*\*weschke@ipk-gatersleben.de  
Insert Length: 559 Std Error: 0.00  
Plate: 10 row: H column: 21  
Seq primer: POB65.

## FEATURES

## source

Location/Qualifiers  
1..559  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="barke"  
/sub\_species="vulgare"  
/db\_xref="GABI:945905"  
/db\_xref="taxon:112509"  
/clone="MPMP2010H2110"  
/tissue\_type="embryo sac"  
/dev\_stage="0-10 DAF (days after flowering)"  
/lab\_host="E. coli, SCS-1/pSE111"  
/clone\_11b="BS1824"  
/note="Vector: pOR30NST (AF074376); Site\_1: SalI; Site\_2: NotI; 0-10 DAF (days after flowering). cDNA synthesis using Bluescript II XR cDNA-library construction kit (Stratagene) with an oligo(dt)-primer containing NotI restriction site and a SalI adapter (Invitrogen). The main library of 21500 clones was rearranged into the sublibrary BS1824 containing 4100 putative expression clones. Note: Due to a cloning artefact caused by the kit, in most cases the SalI site is NOT present, as well as the SalI Adapter used for cloning. To excise the insert, restriction sites upstream SalI should be used (e.g. BamHI). Average insert size is 1 kb. Library generation and sequencing was granted in context of GABI; data are also accessible at https://gabi.rzp.de"

## ORIGIN

Query Match 85.0%; Score 17; DB 7; Length 559;

Best Local Similarity 100.0%; Pred. No. 6.3e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GAAGTGGGCAATCGT 19

DB 540 GAAGTGGGCAATCGT 556

#### RESULT 4

BF627004/c

LOCUS 790 bp mRNA linear EST 22-OCT-2001

DEFINITION HVSMED001P03f Hordeum vulgare seedling shoot EST library

clone HVSMED001P03f, mRNA sequence.

ACCESSION BF627004

VERSION BF627004.2 GI:13088570

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL Unpublished (2001)

COMMENT On Dec 19, 2000 this sequence version replaced gi:11891162.

Contact: Wang RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rtwing@clemson.edu

Total hg bases = 433

Seq primer: AATTAACCTCACTAAGGG

High quality sequence start: 4

High quality sequence stop: 486.

FEATURES

source

1..790

Location/Qualifiers

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

/cultivar="Morex"

/sub\_species="vulgare"

/db\_xref="taxon:112509"

/clone="HVSMED001P03f"

/tissue\_type="Seedling shoot"

/lab\_host="TUC121"

/clone\_1ib="Hordeum vulgare seedling shoot EST library

HVCNDA0002 (Dehydration stress)"

/note="vector: lambdaZAP Site 1: EcoRI; Site 2: XhoI;

Seeds were surface sterilized then germinated under axenic

conditions in the dark at room temperature in covered

with water, nystatin and cefotaxime in covered

crystallization dishes. Five-day old seedlings were

incubated at 90% RH for 24 hr. Shoots were then harvested,

total RNA was prepared, poly(A) RNA was purified, one

primary unamplified cDNA library was made, 600000 plv were

in vivo excised to give Bluescript SK(-) cDNA phagemids.

These steps were performed in the TV Close laboratory at

the University of California, Riverside (Choi, Close,

Fenton). Phagemids were plated and picked at the Clemson

University Genomics Institute (CUGI) (Begum, Palmer,

Friesch, Atkins and Wing). Plasmid DNA preparations, DNA

sequencing and sequence analysis were performed at CUGI

(Wing, Yu, Friesch, Henry, Simmons, Oates, Rambo, Main).

The sequence has been trimmed to remove vector sequence

and contains a minimum of 100 bases of phred value 20 or

above. For more details on library preparation and

sequence analysis see

<http://www.genome.clemson.edu/projects/barley>. To order

this clone see <http://www.genome.clemson.edu/orders> Also

see Close TV, Wing R, Kleinof A, Wise R (2001)

Genetically and physically anchored EST resources for

barley genomics. Barley Genetics Newsletter 31:29-30.

(<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

#### ORIGIN

Query Match 85.0%; Score 17; DB 2; Length 790;

Best Local Similarity 100.0%; Pred. No. 6.7e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GAAGTGGGCAATCGT 19

DB 519 GAAGTGGGCAATCGT 503

#### RESULT 5

CG049330/c

LOCUS 391 bp DNA linear GSS 19-AUG-2003

DEFINITION PU1FQ44TD ZM 0.6.1.0\_KB Zea mays genomic clone ZMMBTC04572H15,

genomic survey sequence.

ACCESSION CG049330

VERSION CG049330.1 GI:33921510

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2003)

Other\_GSSs: PU1FQ44TB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@cigr.org

Seq primer: TP

Class: sheared ends.

Location/Qualifiers

1..391

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBTC04572H15"

/clone\_1ib="ZM 0.6.1.0\_KB"

/note="vector: pCR4-TOPO, Site 1: EcoRI, 0.6-1.0 kb high

COT selected genomic DNA library"

ORIGIN

Query Match

Best Local Similarity

Matches

18; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

QY 1 TCGAAGTGGGCAATCGTG 20

DB 306 TCGAAGTGGGCAATCGTG 287

#### RESULT 6

CL219748/c

LOCUS 452 bp DNA linear GSS 08-JAN-2004

DEFINITION ZMMBBC0447H13f ZMMBBC Zea mays genomic clone ZMMBBC0447H13 5',

genomic survey sequence.

ACCESSION CL219748  
 VERSION CL219748.1 GI:40752422  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 452)  
 Bhatti,A.K., Young,S., Kawchok,S., Keizer,G., Bronzino,A.C., Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.  
 Sequencing of the maize genome at PGIR (2003c)  
 Unpublished (2003)  
 CONTACT Bhatti,A.K.  
 Dr. Joachim Messing's lab  
 The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers University  
 190 Frelinghuysen Road, Piscataway, NJ 08854, USA  
 Tel: 732 445 3801  
 Fax: 732 445 5735  
 Email: bhatti@waksman.rutgers.edu  
 Seq primer: 77  
 Class: BAC ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..452  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBC047H13"  
 /lab\_host="E. coli DH10B"  
 /clone\_1lb="ZMMBC"  
 /note="Vector: pTRABAC1.3; Site\_1: BamHI; Site\_2: BamHI"

ORIGIN  
 Query Match 84.0%; Score 16.8; DB 9; Length 452;  
 Best Local Similarity 90.0%; Pred. No. 7.8e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGTG 20  
 |||||  
 131 TCGAAGTGGTCAATCGTG 112

RESULT 7  
 CC803612 511 bp DNA linear GSS 14-JUL-2003  
 DEFINITION ih17b02.g1 WGS-Zmaysu (DH10b or JM107MA2) Zea mays genomic clone  
 CC803612  
 ih17b02, genomic survey sequence.  
 ACCSSION CC803612  
 VERSION CC803612.1 GI:32628457  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 511)  
 Rabinowicz,P.D., O'Shaughnessy,A.L., Palmer,L., Balija,V., Dedolia,N., Katzenburger,F., King,J., Miller,B., Muller,S., Nascimben,L., Zuberger,T., McCombie,W.R. and Martienssen,R.A.  
 Genomic shotgun sequences from Zea mays (unfiltered)  
 Unpublished (2003)  
 CONTACT W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874  
 Email: mcombie@cshl.org  
 Plate: ih17 row: b column: 02  
 Seq primer: -21M3UnivRev  
 Claes: shotgun

FEATURES  
 source  
 High quality sequence stop: 511.  
 Location/Qualifiers  
 1..511  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="ih17b02"  
 /lab\_host="DH10b or JM107MA2"  
 /note="Organ: immature ears; Site\_1: Xba I; Site\_2: Xba I; The vector was digested with XbaI and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (x/y reads in M13mp19, b/g reads in pUC19). The ligation was transformed into DH10b or JM107MA2."

ORIGIN  
 Query Match 84.0%; Score 16.8; DB 9; Length 511;  
 Best Local Similarity 90.0%; Pred. No. 8e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGTG 20  
 |||||  
 278 TCGAAGTGGTCAATCGTG 297

RESULT 8  
 BU098971 548 bp mRNA linear EST 01-OCT-2003  
 LOCUS BU098971 NIBB Mochli normalized Xenopus early gastrula library  
 DEFINITION Xenopus laevis cDNA clone XL148K21 3', mRNA sequence.  
 BU098971  
 BU098971.1 GI:17601043  
 EST.  
 ACCSSION BU098971  
 VERSION BU098971  
 KEYWORDS Xenopus laevis (African clawed frog)  
 SOURCE Xenopus laevis  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus; Xenopus.  
 1 (bases 1 to 548)  
 Kitayama,A., Terasaka,C., Mochli,M., Ueno,N., Shin-I,T. and Kohara,Y.  
 Expressed genes in X. laevis embryo  
 Unpublished (2001)  
 CONTACT Tadasu Shin-I  
 Center for Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp  
 The information of this clone is available through the following URL.  
 http://xenopus.nibb.ac.jp.  
 Location/Qualifiers  
 1..548  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="XL148K21"  
 /tissue\_type="whole embryo"  
 /dev\_stage="stage 10.5"  
 /clone\_1lb="NIBB Mochli normalized Xenopus early gastrula library"

ORIGIN  
 Query Match 84.0%; Score 16.8; DB 4; Length 548;  
 Best Local Similarity 90.0%; Pred. No. 8e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20  
 DB 474 TCGAAGTGGGCAATCGG 493

RESULT 9  
 B0097494/c 553 bp mRNA linear EST 20-DEC-2002  
 LOCUS phn3d06.y2 Oostertagia oostertagi L3 PAMP1 v1 Oostertagia oostertagi  
 DEFINITION cDNA 5' similar to TR:Q2371 Q2371 T10B10.4 PROTEIN. [1] ; mRNA  
 sequence.

ACCESSION B0097494  
 VERSION B0097494  
 KEYWORDS EST

SOURCE Oostertagia oostertagi  
 ORGANISM Oostertagia oostertagi

REFERENCE 1 (bases 1 to 553)  
 McCarter, J., Clifton, S., Chiappelli, B., Pape, D., Martin, J.,  
 Wyllie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,  
 Bowers, Y., Gibbons, M., Ritzer, B., Bennett, J., Franklin, C.,  
 Tsagaris, V., R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,  
 Underwood, K., Stepcoe, M., Allen, M., Person, B., Swaller, T.,  
 Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,  
 McCann, R., Waterston, R. and Wilson, R.

TITLE The Washington Univ. Nematode EST Project, 1999  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: McCarter JP  
 The Washington Univ. Nematode EST Project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@wustl.wustl.edu  
 The library was constructed by Claire Murphy and Dr. James McCarter  
 at Washington University, St. Louis. DNA Sequencing by: Washington  
 University Genome Sequencing Center  
 High quality sequence stop: 415.

FEATURES  
 source Location/Qualifiers

1..553  
 /organism="Oostertagia oostertagi"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:6317"  
 /dev\_stage="L3"  
 /lab\_host="DH10B"  
 /clone\_lib="Oostertagia oostertagi L3 PAMP1 v1"  
 /note="Vector: PAMP1 (Gibco); Site 1: NotI; Site 2: SalI;  
 \*\*WARNING: Subsequent examination of these samples has  
 revealed the presence of an additional Trichostrongyloidea  
 cattle nematode, Cooperia oncophora. Sequences in this  
 library may derive from either Oostertagia or Cooperia.\*\*  
 The library was constructed by Claire Murphy and Dr. James  
 McCarter at Washington University, St. Louis. The cDNA was  
 made by using Dynabead oligo-dT priming (Dyna1). PCR based  
 library using a modified protocol from the SMART PCR cDNA  
 Synthesis Kit from Clontech. Directionally cloned into the  
 UDG sites of PAMP1. Nematodes were provided by Dr. Louis  
 Gasbarré of the USDA, Beltsville, MD  
 (lgasbarr@nri.barc.usda.gov). Third stage exsheathed  
 larvae were collected from 14 day fecal-ephagnum moss  
 cultures of Oostertagia eggs. The larvae were recovered by  
 overnight passage on a Bertram apparatus, and then  
 cleaned by passage through a 20 micron nylon mesh. The  
 larvae were then subjected to a treatment with 1.25%  
 chlorox to induce excystation. The larvae were washed with  
 5 changes of PBS and then pelleted and snap frozen in  
 liquid nitrogen."

ORIGIN

Query Match 84.0%; Score 16.8; DB 5; Length 553;  
 Best Local Similarity 90.0%; Pred. No. 8.1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20  
 DB 525 TCGAAGTGGGCAATCGTG 506

RESULT 10  
 A2471085/c 607 bp DNA linear GSS 04-OCT-2000  
 LOCUS 1M0285D02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC1M0285D02 R, genomic survey sequence.

ACCESSION A2471085  
 VERSION A2471085  
 KEYWORDS GSS

SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 607)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von  
 Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0285 row: D column: 02  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 607.

FEATURES  
 source Location/Qualifiers

1..607  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="UUGC1M0285D02"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PMD42nv, Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PMD42 (9147321415b) (AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

ORIGIN

Query Match 84.0%; Score 16.8; DB 8; Length 607;  
 Best Local Similarity 90.0%; Pred. No. 8.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCTGTG 20  
|||||  
Db 58 TCGACGTGGGCAAGCGTGTG 39

RESULT 11  
LOCUS BJ460007  
DEFINITION BJ460007 K. Sato unpublished cDNA library, cv. Akashinriki  
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
baak39b17 3', mRNA sequence.  
ACCESSION BJ460007  
VERSION BJ460007.1 GI:21138532  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 608)  
Sato, K., Saiho, D. and Takeda, K.  
Barley EST sequencing project in NIG and Okayama Univ  
JOURNAL Unpublished (2002)  
CONTACT: Tadaeu Shin-1  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers  
1..608  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baak39b17"  
/tissue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_1lb="K. Sato unpublished cDNA library, cv.  
Akashinriki vegetative stage leaves"

ORIGIN  
Query Match 84.0%; Score 16.8; DB 4; Length 608;  
Best Local Similarity 90.0%; Pred. No. 8.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCTGTG 20  
|||||  
Db 142 TCGAAGCGCGCAATCTTGTG 161

RESULT 12  
LOCUS AV910592/c 627 bp mRNA linear EST 18-JAN-2002  
DEFINITION AV910592 K. Sato unpublished cDNA library, cv. Akashinriki  
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
baak1a11 5', mRNA sequence.  
ACCESSION AV910592  
VERSION AV910592.1 GI:18206357  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 627)  
Sato, K., Saiho, D. and Takeda, K.  
Barley EST sequencing project in NIG and Okayama Univ  
JOURNAL Unpublished (2002)  
CONTACT: Tadaeu Shin-1

Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers  
1..627  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baak1a11"  
/tissue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_1lb="K. Sato unpublished cDNA library, cv.  
Akashinriki vegetative stage leaves"

ORIGIN  
Query Match 84.0%; Score 16.8; DB 1; Length 627;  
Best Local Similarity 90.0%; Pred. No. 8.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCTGTG 20  
|||||  
Db 600 TCGAAGCGCGCAATCTTGTG 581

RESULT 13  
LOCUS BJ462359 642 bp mRNA linear EST 23-MAY-2002  
DEFINITION BJ462359 K. Sato unpublished cDNA library, cv. Akashinriki  
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
baak24n23 3', mRNA sequence.  
ACCESSION BJ462359  
VERSION BJ462359.1 GI:21140866  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 642)  
Sato, K., Saiho, D. and Takeda, K.  
Barley EST sequencing project in NIG and Okayama Univ  
JOURNAL Unpublished (2002)  
CONTACT: Tadaeu Shin-1  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers  
1..642  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baak24n23"  
/tissue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_1lb="K. Sato unpublished cDNA library, cv.  
Akashinriki vegetative stage leaves"

ORIGIN  
Query Match 84.0%; Score 16.8; DB 4; Length 642;  
Best Local Similarity 90.0%; Pred. No. 8.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCTGTG 20

Db 13 TCGAAGCGCGCAATCTTG 32

RESULT 14  
LOCUS BJ452459/c

DEFINITION BJ452459 K. Sato unpublished cDNA library, cv. Akashiniriki  
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
baak39b17 5', mRNA sequence.

ACCESSION BJ452459

VERSION BJ452459.1 GI:21131047

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..646

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

/cultivar="Akashiniriki"

/sub\_species="vulgare"

/db\_xref="taxon:112509"

/clone="baak39b17"

/tissue\_type="leaves"

/dev\_stage="vegetative stage"

/clone\_lib="K. Sato unpublished cDNA library, cv.

Akashiniriki vegetative stage leaves"

ORIGIN

Query Match 84.0%; Score 16.8; DB 4; Length 646;

Best Local Similarity 90.0%; Pred. No. 8.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGTGGCGCAATCTTG 20

DB 482 TCGAAGCGCGCAATCTTG 463

RESULT 15

LOCUS AV920519

DEFINITION AV920519 K. Sato unpublished cDNA library, cv. Haruna Nijo

germination shoots Hordeum vulgare subsp. vulgare cDNA clone

bagel4903 3', mRNA sequence.

ACCESSION AV920519

VERSION AV920519.1 GI:18216298

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

1..658

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

/cultivar="Haruna Nijo"

/sub\_species="vulgare"

/db\_xref="taxon:112509"

/clone="bagel4903"

/tissue\_type="shoots"

/dev\_stage="germination"

/clone\_lib="K. Sato unpublished cDNA library, cv. Haruna

Nijo germination shoots"

ORIGIN

Query Match 84.0%; Score 16.8; DB 1; Length 658;

Best Local Similarity 90.0%; Pred. No. 8.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGTGGCGCAATCTTG 20

DB 34 TCGAAGCGCGCAATCTTG 53

Search completed: November 7, 2005, 09:18:42  
Job time : 2008.85 secs

*This Book From*



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:11:16 ; Search time 746.66 seconds  
(without alignments)  
1233.024 Million cell updates/sec

Title: US-10-777-131A-4

Perfect score: 19

Sequence: 1 tggggcttcgctggaagtcg 19

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb ba: \*  
2: gb hcg: \*  
3: gb in: \*  
4: gb om: \*  
5: gb ov: \*  
6: gb pat: \*  
7: gb ph: \*  
8: gb pl: \*  
9: gb pr: \*  
10: gb ro: \*  
11: gb sts: \*  
12: gb sy: \*  
13: gb un: \*  
14: gb vi: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	19	6	AX770106 Sequence
2	19	100.0	1144	6	AX770122 Sequence
3	19	100.0	2406	6	AX770134 Sequence
4	19	100.0	2409	6	BD273957 Sequences
5	19	100.0	2409	6	AX005499 Sequences
6	19	100.0	2451	6	BD273958 Sequences
7	19	100.0	2451	6	AX005501 Sequences
8	19	100.0	2806	6	AX770124 Sequence
9	19	100.0	35336	1	MSGY367
10	19	100.0	103837	1	AE000516_43
11	19	100.0	244800	1	EX842584
12	19	100.0	278492	1	EX248347
13	19	100.0	278492	1	EX248347 Mycobacte
14	17.4	91.6	43836	2	CR847886
15	17.4	91.6	144026	10	AC15304
16	17.4	91.6	175781	10	AC147366
17	17.4	91.6	185769	10	AC129178
18	17.4	91.6	203243	10	AC131184
19	17.4	91.6	219827	2	AC127357
19	17.4	91.6	237551	2	AC094140

20	17	89.5	235148	2	AC113808	AC113808 Rattus no
21	17	89.5	259243	1	AC106670	Rattus no
22	17	89.5	347636	1	EX842653	EX842653 Bdellovib
23	16.4	86.3	2044	10	BC020122	BC020122 Mus muscu
24	16.4	86.3	110000	2	AC095736_1	Continuation (2 of
25	16.4	86.3	110000	2	AC127735_1	Continuation (2 of
26	16.4	86.3	117328	6	C0870156	Sequence
27	16.4	86.3	121813	2	AC147587	AC147587 Orthoborh
28	16.4	86.3	139976	2	AC145482	AC145482 Gallus ga
29	16.4	86.3	167739	2	AC020547	AC020547 Homo sapi
30	16.4	86.3	180000	2	AC102737	AC102737 Mus muscu
31	16.4	86.3	191887	2	CR407563	CR407563 Dantio rer
32	16.4	86.3	236186	2	AC094066	AC094066 Rattus no
33	16.4	86.3	262582	2	AC099145	AC099145 Rattus no
34	16.4	86.3	270991	2	AC136210	AC136210 Gallus ga
35	16.4	86.3	279311	2	AC131439	AC131439 Rattus no
36	16.4	86.3	334051	2	AC123374	AC123374 Rattus no
37	16	84.2	116951	9	AC114788	AC114788 Homo sapi
38	16	84.2	159377	2	AC147522	AC147522 Ocolemur
39	16	84.2	168277	2	AC147934	AC147934 Ocolemur
40	16	84.2	174024	9	AL583827	AL583827 Human DNA
41	16	84.2	185316	2	AC134386	AC134386 Papio anu
42	16	84.2	229063	2	AC147751	AC147751 Ocolemur
43	16	84.2	300162	2	AC149441	AC149441 Meleagris
44	16	84.2	310613	1	AE016875	AE016875 Pseudomon
45	15.8	83.2	201	11	BV179021	BV179021 sqm10356

#### ALIGNMENTS

RESULT 1  
AX770106  
LOCUS AX770106 Sequence 4 from Patent WO03016562. 19 bp DNA linear PAT 02-JUL-2003  
DEFINITION AX770106  
ACCESSION AX770106  
VERSION AX770106.1 GI:32437684  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
1  
AUTHORS  
TITLE  
JOURNAL  
INSTITUT PASTEUR (FR)  
FEATURES  
source  
1..19  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Primer"  
ORIGIN  
Query Match 100.0%; Score 19; DB 6; Length 19;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGGGGTTCGCTGGAAGTCG 19  
1 TGGGGTTCGCTGGAAGTCG 19  
DB 1 TGGGGTTCGCTGGAAGTCG 19  
RESULT 2  
AX770122/c  
LOCUS AX770122 1144 bp DNA linear PAT 02-JUL-2003  
DEFINITION Sequence 20 from Patent WO03016562.  
ACCESSION AX770122  
VERSION AX770122.1 GI:32437700  
KEYWORDS  
SOURCE  
ORGANISM  
Mycobacterium tuberculosis

REFERENCE 1 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

AUTHORS Gicquel,B.

TITLE Compositions and methods for detecting multidrug resistant strains of M. tuberculosis having mutations in genes of the mult family

JOURNAL Patent: WO 03016562-A 20 27-FEB-2003;

INSTITUT PASTEUR (FR)

FEATURES

source Location/Qualifiers

1..1144

/organism="Mycobacterium tuberculosis"

/mol\_type="unassigned DNA"

/db\_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 1144;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

1143 TGGGGTTCGCTGGAAGTGG 1125

Db

RESULT 3

AX770134/c 2406 bp DNA linear PAT 02-JUL-2003

LOCUS Sequence 32 from Patent W003016562.

DEFINITION AX770134

ACCESSION AX770134.1 GI:32437712

VERSION

KEYWORDS

SOURCE

ORGANISM

Mycobacterium tuberculosis

Mycobacterium tuberculosis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 Gicquel,B.

AUTHORS Compositions and methods for detecting multidrug resistant strains of M. tuberculosis having mutations in genes of the mult family

JOURNAL Patent: WO 03016562-A 32 27-FEB-2003;

INSTITUT PASTEUR (FR)

FEATURES

source Location/Qualifiers

1..2406

/organism="Mycobacterium tuberculosis"

/mol\_type="unassigned DNA"

/db\_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2406;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

200 TGGGGTTCGCTGGAAGTGG 182

Db

RESULT 4

BD273957/c 2409 bp DNA linear PAT 17-JUL-2003

LOCUS Sequence nucleic acid from polypeptides exportes from mycobacteries, vector the complement and applications at diagnostic and the prevention from the tuberculose.

DEFINITION BD273957

ACCESSION BD273957.1 GI:33083725

VERSION

KEYWORDS

SOURCE

ORGANISM

Mycobacterium tuberculosis

Mycobacterium tuberculosis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 (bases 1 to 2409)

AUTHORS Gicquel,B., Portnoi,D., Lam,E., Pelletier,V., Guigueno,A. and Salmoniere,Y.G.D.L.

TITLE Sequences nucleic acid from polypeptides exportes from mycobacteries, vector the complement and applications at diagnostic and the prevention from the tuberculose

JOURNAL Patent: JP 2002534956-A 181 22-OCT-2002;

INSTITUT PASTEUR

COMMENT OS Mycobacterium tuberculosis

PN JP 2002534956-A/181

PD 22-OCT-2002

PF 14-AUG-1998 JP 2000509849

PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI

BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELLETIER, PI

AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE

PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,

PC C07K16/12,

PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC

G01N33/53//

PC (C12N15/09,C12R1.32),C12N15/00,(C12N15/00,C12R1.32) CC SEQ

ID No. 36D

FT Key Location/Qualifiers

(1)..(2406).

CDS

FEATURES

source Location/Qualifiers

1..2409

/organism="Mycobacterium tuberculosis"

/mol\_type="genomic DNA"

/db\_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2409;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

200 TGGGGTTCGCTGGAAGTGG 182

Db

RESULT 5

AX005499/c 2409 bp DNA linear PAT 24-AUG-2000

LOCUS Sequence 618 from Patent W0909186.

DEFINITION AX005499

ACCESSION AX005499.1 GI:9928645

VERSION

KEYWORDS

SOURCE

ORGANISM

Mycobacterium tuberculosis

Mycobacterium tuberculosis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 Portnoi,D. and Guigueno,A.

AUTHORS Polypeptide nucleic sequences exported from mycobacteria, vectors comprising same and uses for diagnosing and preventing tuberculosis

JOURNAL Patent: WO 9909186-A 618 25-FEB-1999;

PORTNOI DENIS (FR); GUIGUENO AGNES (FR)

FEATURES

source Location/Qualifiers

1..2409

/organism="Mycobacterium tuberculosis"

/mol\_type="unassigned DNA"

/db\_xref="taxon:1773"

1..2409

/note="unnamed protein product; SEQ ID NO 36D"

/codon\_start=1

/transl\_table=1

/protein\_id="CAC05103.1"

/db\_xref="GI:9928646"

/translation="MTALQLGMAALARYTSAGVAGIGMALTPVSAAPHALAGRSP

TPFVORIDQVTPVVTSSSEPHVTSGVTNTGDRPVQDVNRLEHAAVTSSTALR

TSIAGGDDQGPADDFLTVAPELRGAGFTLSAPRSLRPSIAVNOGIVPVLVN

VNGRPDYGPAPRLNARFLRVVGVPPQADTFPSAVAPETTPATWTMTMMLARPR

LPAPGQTVFVRLVDDLLANSLANGRLDITLISAAFPATREVPDQAVGRALCLAI

PDLLITIVNMTGIVVSDSPDGAALPPTPTHTPGTGOAASWLDRLRTVHTCYV  
PLPFAQADLDLORVNDPRLSALITISPADIVDRILVDSSTEGATVLPDGPVGRAIN  
LSTHGNTVAVAADSPREBOGSSOISALIPAPRSLSPRVAAAPBDPAVGAALA  
AAGNPTVPTLIDPSLFVRIABESITARODALGAMLRSEPNAPRTQILVPPASW  
SLASDDAQVILITALTATIRSGLAVERPLPAVYADAAATEPEEPNAPGSAARGPND  
ITTOIGGVAVLMLKLTSLITIDRTGLTGVOTLARIDMLALSCQSJPPDRNLQAO  
ORLAVGKIDLDLFGAVTIVNGGSYTTATSHSPPLALHNGLAVERLQVDPAPG  
MTVADVGOIBLPGVLPPLRVPIEVNFTORVAVDSLRTPDGVALGEPVRLGVSNAVG  
KVLPAITLSAAAVLVTLAAGRILMHRFRQGPBRADRDRLPTGKHAPORAVASNDDE  
KHRV"

## ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2409;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGGGTTGCTGGAAGTGG 19  
Db 200 TGGGGTTGCTGGAAGTGG 182

## RESULT 6

BD273958/c 2451 bp DNA linear PAT 17-JUL-2003

DEFINITION Sequences nucleic acid from polypeptides exported from mycobacteries, vector the complement and applications at diagnostic and the prevention from the tuberculose.

ACCESSION BD273958  
VERSION BD273958.1 GI:33083726

KEYWORDS JP 2002534956-A/182.

SOURCE Mycobacterium tuberculosis

ORGANISM Mycobacterium tuberculosis

REFERENCE Corynebacterineae; Actinobacteridae; Actinomycetales; Bacteria; Actinobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

AUTHORS 1 (bases 1 to 2451)

TITLE Gicquel,B., Portnoi,D., Lim,E., Pellicic,V., Guisgueno,A. and Salmoniere,Y.G.D.L.

Sequences nucleic acid from polypeptides exportees from mycobacteries, vector the complement and applications at diagnostic and the prevention from the tuberculose

JOURNAL Patent: JP 2002534956-A 182 22-OCT-2002;

COMMENT INSTITUTEUR  
OS Mycobacterium tuberculosis

PN JP 2002534956-A/182

PD 22-OCT-2002

PF 14-AUG-1998 JP 2000509849

PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI

BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI

AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE

PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,

PC C07K16/12,

PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC

PC G01N33/53//

PC (C12N15/09,C12R1/32),C12N15/00,(C12N15/00,C12R1/32) CC SEQ

ID No. 36F

FI Key

FT CDS Location/Qualifiers

1. 2451 Location/Qualifiers

/organism="Mycobacterium tuberculosis"

/mol\_type="genomic DNA"

/db\_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2451;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGGGTTGCTGGAAGTGG 19

Db 242 TGGGGTTGCTGGAAGTGG 224

## RESULT 7

AX005501/c

LOCUS

DEFINITION

AX005501

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

CDS

AX005501

Sequence 620 from Patent WO909186.

AX005501

GI:9928647

Myobacterium tuberculosis

Bacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium

tuberculosis complex.

Portnoi,D. and Guisgueno,A.

Polypeptide nucleic sequences exported from mycobacteria, vectors

comprising same and uses for diagnosing and preventing tuberculosis

Patent: WO 9909186-A 620 25-FEB-1999;

PORTNOI DENIS (FR); GUIGUENO AGNES (FR)

Location/Qualifiers

1. 2451

/organism="Mycobacterium tuberculosis"

/mol\_type="unassigned DNA"

/db\_xref="taxon:1773"

<4>.2451

/note="unamed protein product; SEQ ID NO 36F"

/codon\_start=1

/transl\_table=1

/protein\_id="CAC05104.1"

/db\_xref="GI:9928648"

/translation="LSTGSAQRSGAVATLQGLAALATVSAIGVAGMALTVP

SAAPHALGSPSPFPVOVRIDQVTPVVTSSSPHVTGVTNMGPRPDVWRL

EHAAVTSSTLRSLSDGCTQYQPADFLVABEDRQGAAGTSLAPLSRPSL

AVNPGIPLVAVNAGTDFGAPRLNARLPLVAVGPPQATDFGSAVAPETAV

WITMLPLADRPAPGAPGTPVRLVDDLSLANSGLRLDILLSAEPATKREVD

PDGAVGRALCIAIDPDLITVNTGTVSDSPDGAOLPPTPTHTPGTGOAASWT

DRLETVHRTCTVPLPAQADLDLORVNDPRLSAIATISPADIVDRILDVSSRGAT

VLPPGTLGRALINLSTHGNTVVAADSPREBOGSSOISALIPAPRSLSPRV

AAPRPVGAALAAAGTNPVPTVLTDSLPRIABESITARODALGAMLRSEPN

APRTQILVPPASMSLASDDAVIITALTATIRSGLAVERPLPAVYADAAATEPEEP

GAYSAAGRENDITTOIGQVAVLMLKLTSLITIDRTGLTGVOTLARIDMLALS

QSLPPDRNGLAQORLAVGKIDLDLFGAVTIVNGGSYTTATSHSPPLALHNGLA

PIRVLOVDAPGKTVADVGOIELPGLVPLRVPIEVNFTORVAVDSLRTPDGV

EPVRLSVNSAVGKVLPAITLSAAAVLVTLAAGRILMHRFRQGPBRADRDRLPTGK

APORRAVASRDEKRV"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2451;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGGGTTGCTGGAAGTGG 19

Db 242 TGGGGTTGCTGGAAGTGG 224

## RESULT 8

AX770124/c

LOCUS

DEFINITION

AX770124

ACCESSION

AX770124

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Sequence 22 from Patent WO03016562.

AX770124.1 GI:32437702

Myobacterium tuberculosis

Bacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex.

Gicquel,B.

Compositions and methods for detecting multidrug resistant strains of M. tuberculosis having mutations in genes of the mutT family

Patent: WO 03016562-A 22 27-FEB-2003;

```

FEATURES             INSTTUT PASTEUR (FR)
    source            Location/Qualifiers
                        1..2806
                        /organism="Mycobacterium tuberculosis"
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:1773"
ORIGIN
Query Match          100.0%; Score 19; DB 6; Length 2806;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
        |||||||
Db       400 TGGGGTTGCTGGAAGTGG 382

RESULT 9
MSGY367          3536 bp   DNA   linear   BCT 03-DEC-1996
LOCUS            Mycobacterium tuberculosis sequence from clone y367.
ACCESSION        AD000008
VERSION          AD000008.1 GI:1702971
KEYWORDS
SOURCE
ORGANISM
    Mycobacterium tuberculosis
    Mycobacterium tuberculosis
    Bacteria; Actinobacteria; Actinomycetales;
    Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
    tuberculosis complex.
    1 (bases 1 to 3536)
REFERENCE
    Du, L.
    Direct Submission
    Submitted (11-OCT-1996) L. Du, Genome Therapeutics Corporation, 100
    Beaver Street, Waltham, MA, USA, 02154 du@cric.com
COMMENT
    FEATURES
        source
            1..3536
            Location/Qualifiers
                /organism="Mycobacterium tuberculosis"
                /mol_type="genomic DNA"
                /db_xref="taxon:1773"
                /clone="y367"
ORIGIN

Query Match          100.0%; Score 19; DB 1; Length 3536;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
        |||||||
Db       13574 TGGGGTTGCTGGAAGTGG 13592

RESULT 10
AE000516.43/c
WPCOMMENT
Sequence split into 44 fragments LOCUS AE000516 Accession AE000516
Fragment Name      Begin      End
AE000516_00        1      110000
AE000516_01        100001  210000
AE000516_02        200001  310000
AE000516_03        300001  410000
AE000516_04        400001  510000
AE000516_05        500001  610000
AE000516_06        600001  710000
AE000516_07        700001  810000
AE000516_08        800001  910000
AE000516_09        900001  1010000
AE000516_10       1000001  1110000
AE000516_11       1100001  1210000
AE000516_12       1200001  1310000
AE000516_13       1300001  1410000
AE000516_14       1400001  1510000
AE000516_15       1500001  1610000

```

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCES	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCES
AE000516_16		1600001	1710000									
AE000516_17		1700001	1810000									
AE000516_18		1800001	1910000									
AE000516_19		1900001	2010000									
AE000516_20		2000001	2110000									
AE000516_21		2100001	2210000									
AE000516_22		2200001	2310000									
AE000516_23		2300001	2410000									
AE000516_24		2400001	2510000									
AE000516_25		2500001	2610000									
AE000516_26		2600001	2710000									
AE000516_27		2700001	2810000									
AE000516_28		2800001	2910000									
AE000516_29		2900001	3010000									
AE000516_30		3000001	3110000									
AE000516_31		3100001	3210000									
AE000516_32		3200001	3310000									
AE000516_33		3300001	3410000									
AE000516_34		3400001	3510000									
AE000516_35		3500001	3610000									
AE000516_36		3600001	3710000									
AE000516_37		3700001	3810000									
AE000516_38		3800001	3910000									
AE000516_39		3900001	4010000									
AE000516_40		4000001	4110000									
AE000516_41		4100001	4210000									
AE000516_42		4200001	4310000									
AE000516_43		4300001	4403837									
Continuation (44 of 44) of AE000516 from base 4300001 (AE000516 Mycobacterium tuberculosis)												
Query Match		100.0%	Score 19;	DB 1;	Length 103837;							
Best Local Similarity		100.0%	Pred. No. 29;									
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;								
QY	1											
Db	86697	TCGGGTTCCGTGGAATG	19	TCGGGTTCCGTGGAATG	86679							
RESULT 11												
LOCUS	BX842584	244800	bp	DNA	linear	BCT 10-JUN-2004						
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 13/13											
ACCESSION	BX842584	AL022126	AL022076	AL022120	AL022121	AL123456	280343	283864				
VERSION	BX842584.1	GI:38490370										
KEYWORDS	complete genome.											

score is 0.900."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"  
/protein\_id="CAA18045.1"  
/db\_xref="GI:2960147"  
/translation="MGCRVAVLMMHASPISGAGVLYFYFVLPREMLGDGHSCTGIL  
RIATAYGLALPVPFTLTKRPELCTPOLASMTISMAHYALGALVGRATISS  
VWLSIDAAGQVLFYIGAAALAVLGFYFGFLTSVAEIPPPPLPKPKKQRLRR  
KKTAAIGDSEAAEAEETLAAQDEAEVAEPESIESGGESEATREAPAAETAL  
TAEERGLRRNRRTGKTSRRRRTRSGVAKVDE"  
2735..2977  
/gene="cut5a"  
/locus\_tag="RV3724A"  
2735..2977  
/gene="cut5a"  
/locus\_tag="RV3724A"  
/EC\_number="3.1.1.-"  
/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE  
STRUCTURE OF PLANT CUTICLE)."  
/note="RV3724A, (MTV025.072), len: 80 aa. Probable cut5a,  
truncated cutinase precursor (EC 3.1.1.-), similar to  
N-terminal end of others e.g. Q9KX87 SERINE ESTERASE  
CUTINASE from *Mycobacterium avium* (220 aa), FASTA scores:  
opt: 202, E(): 1.5e-06, (56.45% identity in 62 aa  
overlap); Q9X809|RV2D-RV1758 PROTEIN (FRAGMENT) from  
*Mycobacterium bovis* BCG (143 aa), FASTA scores: opt: 200,  
E(): 1.5e-06, (61.4% identity in 57 aa overlap); and  
Q00288|CUT1\_BOTCI|CUT1A CUTINASE PRECURSOR from *Botrytis  
cinerea* (*Botryotinia fuckeliana*) (202 aa), FASTA scores:  
opt: 109, E(): 2.2, (40.4% identity in 52 aa overlap).  
Also highly similar to others from *Mycobacterium  
tuberculosis* e.g.  
006318|CUT1\_MYCTU|RV3451|MT3557|MTCY31E12.04 PROBABLE  
CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 189, E():  
1.2e-05, (58.0% identity in 50 aa overlap);  
Q506641|CUT2\_MYCTU|RV2301|MT2358|MTCY339.08C PROBABLE  
CUTINASE PRECURSOR (219 aa), FASTA scores: opt: 172, E():  
0.00015, (59.2% identity in 49 aa overlap);  
006793|RV1758|MTCY28.24|Z95890 HYPOTHETICAL 17.9 KDA  
PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29,  
(57.2% identity in 166 aa overlap);  
00619|RV3452|MTY13E12.05; and U00015\_11 from *Mycobacterium  
leprae*. BELONGS TO THE CUTINASE FAMILY. Rest of cutinase  
ORF continues as RV3724B|CUT5B, frameshifting could occur  
near position 4169666. Sequence has been checked but no  
errors found."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE CUTINASE PRECURSOR [FIRST PART] CUT5A"  
/protein\_id="CAE55633.1"  
/db\_xref="GI:38490372"  
/translation="MDVIYMRRLAVVAGTAAVTTGGLSAHVPMVSAEPCDVEVVV  
FARGGEPGIGSVGLFVDALRPFGRQRYTGRAR"  
2874..3437  
/gene="cut5b"  
/locus\_tag="RV3724B"  
2874..3437  
/gene="cut5b"  
/locus\_tag="RV3724B"  
/EC\_number="3.1.1.-"  
/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE  
STRUCTURE OF PLANT CUTICLE)."  
/note="RV3724B, (MTV025.072), len: 187 aa. Probable cut5b,  
truncated cutinase (EC 3.1.1.-), similar to C-terminal end  
of others e.g. Q9X809|RV2D-RV1758 PROTEIN (FRAGMENT) from  
*Mycobacterium bovis* BCG (143 aa) FASTA scores: opt: 335,  
E(): 3.4e-12, (53.25% identity in 92 aa overlap); Q9KX87  
SERINE ESTERASE CUTINASE from *Mycobacterium avium* (220  
aa), FASTA scores: opt: 251, E(): 2.5e-07, (44.05%  
identity in 168 aa overlap). Also similar to proteins from  
*HYPOTHETICAL 17.9 KDA PROTEIN* (174 aa), FASTA scores: opt:

641, E(): 2.5e-29, (57.25% identity in 166 aa overlap);  
006319|RV3452|MYCY13B12.05 HYPOTHETICAL 23.1 KDA PROTEIN  
(1226 aa), FASTA scores: opt: 385, E(): 7.5e-15, (46.65%  
identity in 165 aa overlap);  
006318|CUT3\_MYCTU|RV3451|MT1557|MYCY13B12.04 PROBABLE  
CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 307, E():  
1.9e-10, (40.7% identity in 167 aa overlap);  
010837|CUT1\_MYCTU|RV1984c|MT2037|MYCY39.35 PROBABLE  
CUTINASE PRECURSOR (217 aa), FASTA scores: opt: 261, E():  
6.7e-08, (50.9% identity in 169 aa overlap); etc; and  
U00015\_11 from Mycobacterium lepra. 5'-end of gene is  
RV3724A|CUT5A; frameshifting may occur near position  
4169668. Tbpase score is 0.918."  
/codon\_start=1  
/transl\_table=11  
/product="PROBABLE CUTINASE [SECOND PART] CUT5B"  
/protein\_id="CAE55633.1"  
/db\_xref="GI:38490373"  
/translation="MAPGSHLVLAASEDCSSTHCYSQVGAASLGVYAVNPASNDPAS  
SDPKTVIDGIRDAAGSHIQSMAMSCPTROYGAGYSGAAVAGYVTSVAVPAPVQA  
VPAPAPVAVNHVAIVTLFGAPSAQFLQYGAIPAIQPLVQPKTLQCADGDSICGD  
GNSPVAHLVAVNGVVGAGAPFAASRL"

Query Match 100.0%; Score 19; DB 1; Length 244800;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGCTTCGTCGGAAGTGG 19  
|||||  
Db 227659 TGGGCTTCGTCGGAAGTGG 227641

RESULT 12  
LOCUS BX248347/c 278492 bp DNA linear BCT 06-MAY-2004  
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment  
14/14.  
ACCESSION BX248347 BX248333  
VERSION BX248347.1 GI:31620456  
KEYWORDS complete genome.  
SOURCE Mycobacterium bovis AF2122/97  
ORGANISM Mycobacterium bovis AF2122/97  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
REFERENCE 1  
AUTHORS Garnier,T., Eiglemaier,K., Camus,J.-C., Medina,N., Mansoor,H.,  
Pryor,M., Duchoy,S., Grondin,S., Lacroix,C., Konsempe,C., Simon,S.,  
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,  
Wheeler,P.R., Parkhill,J., Barrrell,B.G., Cole,S.T., Gordon,S.V. and  
Hewison,G.  
TITLE The complete genome sequence of Mycobacterium bovis  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 278492)  
Garnier,T.  
REFERENCE Direct Submision  
JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire  
Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex  
15, France. e-mail:fgarnier@pasteur.fr Submitted on behalf of the  
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary  
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,  
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut  
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28  
rue du Docteur Roux, 75724 Paris Cedex 15, France

FEATURES  
source  
1..278492  
/organism="Mycobacterium bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:233413"  
complement(240..704)  
gene

CDS

/locus\_tag="MB3713c"  
complement(240..704)  
/locus\_tag="MB3713c"  
/note="MB3713c", len: 154 aa. Equivalent to RV3688c,  
len: 154 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 154 aa overlap). Hypothetical protein,  
similar to other bacterial hypothetical proteins e.g.  
O9X341|SCHE6.02c from Streptomyces coelicolor (154 aa),  
FASTA scores: opt: 425, E(): 3.4e-19, (46.1% identity in  
154 aa overlap); O9WZF4|TM0690 from Thermotoga maritima  
(149 aa), FASTA scores: opt: 326, E(): 3.4e-13, (40.4%  
identity in 151 aa overlap); O9PHU3|CU0573 from  
Campylobacter jejuni (147 aa), FASTA scores: opt: 290,  
E(): 5.1e-11, (36.4% identity in 151 aa overlap); etc.  
Also some similarity to upstream  
O69654|RV3686c|MTV025.034c CONSERVED HYPOTHETICAL PROTEIN  
from Mycobacterium tuberculosis.

gene

CDS

/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD95899.1"  
/db\_xref="GI:31620457"  
/db\_xref="UniProt/TREMBL:Q7TVX7"  
/translation="MAELKSQRLSDLTQAMKTQDKLTATRTMTLAIQTSEVSKQA  
RELSDDEVITKVLARESRKRGRABRIYONGEALAEAHARARIIDEVLPPLTEGEL  
ADVADTALAEVABEILGHRPSMKOMGLVMAKATVIAAGKADGARSAAVKERL"  
704..2059  
/locus\_tag="MB3714"  
704..2059  
/locus\_tag="MB3714"  
/note="MB3714", len: 451 aa. Equivalent to RV3689, len:  
451 aa, from Mycobacterium tuberculosis strain H37Rv,  
(99.8% identity in 451 aa overlap). Probable conserved  
transmembrane protein, with proline rich N-terminus,  
similar to O9KTM6|SC833.17 PUTATIVE INTEGRAL MEMBRANE  
PROTEIN from Streptomyces coelicolor (462 aa) FASTA  
scores: opt: 730, E(): 2.7e-21, (38.1% identity in 412 aa  
overlap)."  
/codon\_start=1  
/transl\_table=11  
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"  
/protein\_id="CAD95900.1"  
/db\_xref="GI:31620458"  
/db\_xref="UniProt/TREMBL:Q7TVX6"  
/translation="MHKRYAPORPKPDRTYIEKCTDRQDGHDERQLRPVAMP  
PGYVAPPPVAVAGYAPGYPVPAATPPGPGGAGAPSPGPPGPGYVAPAPPG  
CGPPPGPPGYPPLAPAVKKGIIPLRLTSLDITNGAVCYIRAPAKTGLGTAM  
VMTLTIISLVALGPPMTAFEDITGEBDELTGAVGMSAFGASLVMSVAGVLSA  
GMLTVIVGKAVFGSPITVGRAMAVRGLLAFGLALEAGVAVLDAVAVILSGVA  
AAANBAALALGFDLLLVGVSILVLYLVLLFAPVLYLERLPIVEATRTSPALVRHG  
FMRVIGRLITVLVVGVGNAIAAPFMVIGRIYAVTASDGSVTRVETLGCATLSAIGVT  
IGQIVTAPFSGVAVLLYTDRIIRAEARDVLTQGLGAPGAPGAPVSTNLMWLRP  
F"

gene

CDS

2086..2739  
/locus\_tag="MB3715"  
2086..2739  
/locus\_tag="MB3715"  
/note="MB3715", len: 217 aa. Equivalent to RV3690, len:  
217 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 217 aa overlap). Probable conserved  
membrane protein, similar to O9KY45|SC83.18 PUTATIVE  
INTEGRAL MEMBRANE PROTEIN from Streptomyces coelicolor  
(231 aa), FASTA scores: opt: 419, E(): 1.5e-19, (36.0%  
identity in 211 aa overlap). Equivalent to AAK48159 from  
Mycobacterium tuberculosis strain CDC1551 (233 aa) but  
shorter 16 aa."

/codon\_start=1  
/transl\_table=11  
/product="PROBABLE CONSERVED MEMBRANE PROTEIN"  
/protein\_id="CAD95901.1"  
/db\_xref="GI:31620459"  
/db\_xref="UniProt/TREMBL:Q7TVX5"  
/translation="WPSIDIREAAHQARELDRKPIYKDSLTKELTDWIDELQYRI

LEKSGISGNGMTTIVLLIMIAVTAALVQIARSTPRNRGCDUQDPAGUQTAACHR  
STAEVYAEAGGMAAIBRRLOAVARELEETGLMNPAGACTIADSGEVLPLABAE  
LTOATATNDVYTGERRPGTQAYQMTADLDHLNRSFAVVAIVQHPAFDSMAQVR"  
2865. .3866  
/locus\_tag="Mb3716"  
2865. .3866  
/locus\_tag="Mb3716"  
/locus\_tag="Mb3716"  
/note="Mb3716, -, len: 333 aa. Equivalent to Rv3691, len: 333 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 333 aa overlap). Conserved hypothetical protein, similar to Q9KW41|SC33.19 PUTATIVE HYPOTHECAL PROTEIN from Streptomyces coelicolor (387 aa) FASTA scores: opt: 481, E(): 6e-23, .36 6% identity in 358 aa overlap). Equivalent to AAK8160 from Mycobacterium tuberculosis strain CDC1551 (361 aa) but shorter 48 aa."  
/codon\_start=1  
/transl\_table=1  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD95902.1"  
/db\_xref="GI:31620460"  
/db\_xref="UniProt/TREMBL:Q7TVX4"  
/translation="MAPASTSTSGHATATLGNHGEVVVADSVADVEAARPSDILLVAQYQVNDALDRKAPDLGADLVLSGALVPRFAGRTITVAASFPNSQPCITRBNRSGVSDSTYQATNGDLVTSXGVALVPRAGRTITVAASFPNSQPCITRAGNALRIGLNGRNPRIYVAAPDHIEGMSRSSLSIDIPENHMTIWMQLVYLVAALMKGRIRGPIYAAELPVITRASETVEGGRGLYRSRRARDRAADLRITVLOLRPRIGVGAGAPAAVAVTTIAQRSKADPPFAVHLFGAPATNDLILQIARALDIERQVTHS"  
3863. .4939  
/gene="moxr2"  
3863. .4939  
/locus\_tag="Mb3717"  
3863. .4939  
/gene="moxr2"  
/locus\_tag="Mb3717"  
/note="Mb3717, moxr2, len: 358 aa. Equivalent to Rv3692, len: 358 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 358 aa overlap). Probable moxr2, methanol dehydrogenase regulatory protein, highly similar (generally longer at N-terminus) to Q9KW43|SC33.20 PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (329 aa), FASTA scores: opt: 1523, E(): 4.2e-74, (70.9% identity in 330 aa overlap); .092538|SC98B.21c PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (332 aa) FASTA scores: opt: 1008, E(): 1.1e-46, (50.8% identity in 313 aa overlap); .Q9U267|MOXR-3|PAB0848 METHANOL DEHYDROGENASE REGULATORY PROTEIN from Pyrococcus abyssi (314 aa), FASTA scores: opt: 989, E(): 1.1e-45, (50.65% identity in 302 aa overlap); .Q9AAN1|CO0566 MOXR PROTEIN from Caulobacter crescentus (323 aa), FASTA scores: opt: 988, E(): 1.3e-45, (52.3% identity in 306 aa overlap); etc. Also similar to O53170|MY007.26|MOXR|rv1479 from Mycobacterium tuberculosis (377 aa); and .O07392|AF002133.6|MOXR from Mycobacterium avium (309 aa). Also high similarity with several hypothetical bacterial proteins."  
/codon\_start=1  
/transl\_table=1  
/product="PUTATIVE METHANOL DEHYDROGENASE TRANSCRIPTIONAL REGULATORY PROTEIN MOXR2"  
/protein\_id="CAD95903.1"  
/db\_xref="GI:31620461"  
/db\_xref="UniProt/TREMBL:Q7TVX3"  
/translation="WTQASNPQAPPTQPCAEI,PGVPPQAGAPTAAPSGPGRHRAE,AEASDALILAAEAVAKVQGDQISGLVALLCRGAVLLIEGVGAKTLIVAMSAALQLEFKVQVFPDLPMDGVTSGLYDARTRAEFVRRPQPTVNTLLDDEIRRPKPTQALIAEMERQVSVSGEKPLPNPRTIVATQNPILVESTQVLO,PAQGLDRIKNTVTLPARDEITILIDHHAQFDPDRDISAINPAGAEIAAGAEARHVLVANEVLGIVYIVGATRSAPPLQIQEVSFPGATALLGTIRSAWMLSGRDYVTPDDVKAMARPTLHRWMLPRAELGASPPVGLPDLIDSIASVYPR"  
5073. .6395  
/locus\_tag="Mb3718"  
5073. .6395  
/locus\_tag="Mb3718"  
/locus\_tag="Mb3718"  
/note="Mb3718, -, len: 440 aa. Equivalent to Rv3693, len:

440 aa (alternative start at 41910), from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 440 aa overlap). Possible conserved membrane protein, similar to O9K1W2|SC833.21 PUTATIVE LIPOPROTEIN from Streptomyces coelicolor (436 aa), FASTA scores: opt: 875, E(1): 3.3e-46, (56.25% identity in 448 aa overlap); O9AN0|CC0567 HYPOTHEICAL PROTEIN from Caulobacter crescentus (437 aa), FASTA scores: opt: 355, E(1): 2.3e-14, (30.9% identity in 450 aa overlap); P7J2J3|SLR2013 HYPOTHEICAL 46.5 KDA Mammalia from Synecocystis sp. strain PCC 6803 (435 aa), FASTA scores: opt: 340, E(1): 1.9e-13, (29.7% identity in 438 aa overlap); etc. Equivalent to AAK48162 from Mycobacterium tuberculosis strain CDC1551 (475 aa) but shorter 35 aa. Also similar to other hypothetical proteins from Mycobacterium tuberculosis; MTV014\_7; MTV007\_27; and

100.0%; Score 19; DB 1; Length 278492;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGTGAAGTGG 19  
|||||  
Db 261352 TGGGGTTGCTGTGAAGTGG 261334

RESULT 13  
CR847886/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

CR847886 43836 bp DNA linear HTG 06-OCT-2004  
Mus musculus chromosome X clone W11-2134P12, WORKING DRAFT  
SEQUENCE, 2 unordered pieces.  
CR847886  
CR847886.1 GI:53850224  
HTG; HTGS PHASE1; HTGS\_ACTIVEPIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 43836)  
Grafham, D.  
Direct Submission  
Submitted (05-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk  
Sequence from the Mouse Genome Sequencing Consortium whole genome  
shotgun may have been used to confirm this sequence. Sequence data  
from the whole genome shotgun alone has only been used where it has  
a phred quality of at least 30.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: FM2134P12  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 43356 bases at least Q40  
Consensus quality: 43507 bases at least Q30  
Consensus quality: 43620 bases at least Q20  
Insert size: 43736; sum-of-contigs  
Insert size: 51468; 2.9% error; agarose-fp  
Quality coverage: 6.2ix in Q20 bases; sum-of-contigs Quality  
coverage: 6.12x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.



FEATURES

\* 1 10908: contig of 10908 bp in length  
\* 11009 11008: gap of 100 bp  
\* 11009 43836: contig of 32828 bp in length.  
Location/Qualifiers  
1. 43836  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="W11-2134P12"  
/clone\_11b="Whitehead\_mouse\_foamid"  
1. 10908  
/note="assembly\_fragment:00146  
fragment\_chain:1"  
11009. 43836  
/note="assembly\_fragment:00240  
fragment\_chain:1  
clone\_end:SP6  
vector\_side:right"

ORIGIN

Query Match 91.6%; Score 17.4; DB 2; Length 43836;  
Best Local Similarity 94.7%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TGGGGTTCGCTGAGTGG 19  
DB 10688 TGGGGTTCGCTGAGTGG 10670

RESULT 14  
AC115304 144026 bp DNA linear ROD 05-NOV-2003  
LOCUS Mus musculus BAC clone RP24-156K1 from 15, complete sequence.  
ACCESSION AC115304  
VERSION AC115304.3 GI:21844659  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 144026)  
TITLE Vanbrunt, A., Haakenson, W. and Boyer, E.  
JOURNAL The sequence of Mus musculus BAC clone RP24-156K1  
REFERENCE 2 (bases 1 to 144026)  
TITLE Wilson, R.  
JOURNAL Sequencing of Mus musculus  
REFERENCE 3 (bases 1 to 144026)  
TITLE Unpublished (2001)  
AUTHORS McPherson, J.D. and Waterston, R.H.  
JOURNAL Direct Submission  
TITLE Submitted (16-MAR-2002) Genome Sequencing Center, 4444 Forest Park  
JOURNAL Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 144026)  
TITLE Unpublished (2001)  
AUTHORS McPherson, J.D. and Waterston, R.H.  
JOURNAL Direct Submission  
TITLE Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
JOURNAL Parkway, St. Louis, MO 63108, USA  
REFERENCE 5 (bases 1 to 144026)  
TITLE Unpublished (2001)  
AUTHORS McPherson, J.D. and Waterston, R.H.  
JOURNAL Direct Submission  
TITLE Submitted (16-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
JOURNAL Parkway, St. Louis, MO 63108, USA  
REFERENCE 6 (bases 1 to 144026)  
TITLE Unpublished (2001)  
AUTHORS Wilson, R.  
JOURNAL Direct Submission  
TITLE Submitted (05-NOV-2003) Department of Genetics, Washington  
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT On Jul 16, 2002 this sequence version replaced gi:21040012.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
Center project name: M\_BB0156K01  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:  
The RPCT-24 BAC library has been constructed by Pieter de Jong and coworkers (<http://www.choxi.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.choxi.org>

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone.

FEATURES

source

1. 144026  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="15"  
/map="15"  
/clone="RP24-156K1"  
/clone\_11b="RPCT-24"  
1748..1887  
/rpt\_family="Alu"  
4935..5023  
/rpt\_family="B2"  
6425..6472  
/rpt\_family="Alu"  
6494..6695  
/rpt\_family="B2"  
7330..7564  
/rpt\_family="CR1"  
7621..7734  
/rpt\_family="Alu"  
7828..7967  
/rpt\_family="B4"  
8812..8948  
/rpt\_family="B4"  
10054..10188  
/rpt\_family="Alu"  
10289..10472  
/rpt\_family="B2"  
11445..11563  
/rpt\_family="Alu"  
11657..11718  
/rpt\_family="Alu"  
11662..11743  
/rpt\_family="B4"  
11933..12035  
/rpt\_family="Alu"  
12368..12447  
/rpt\_family="L1"  
12828..12977  
repeat\_region



```
/rpt_family="B4"
repeat_region 1303T .13193
/rpt_family="B2"
repeat_region 1342T .13627
/rpt_family="MER1_type"
repeat_region 14279 .14400
/rpt_family="B4"
repeat_region 14547 .14679
/rpt_family="B2"
repeat_region 14716 .15006
/rpt_family="MALR"
repeat_region 16277 .16411
/rpt_family="Alu"
repeat_region 16416 .16494
/rpt_family="ID"
repeat_region 17682 .17927
/rpt_family="L1"
repeat_region 17929 .18103
/rpt_family="B2"
repeat_region 18106 .18223
/rpt_family="L1"
repeat_region 18293 .18447
/rpt_family="MER2_type"
repeat_region 18448 .18629
/rpt_family="B2"
complement(18551..18623)
/product="crRNA-Ser"
/note="Likely pseudogene (HMM Sc=39.59 / Sec struct
Sc=-11.21)"
18630..18678
/rpt_family="MER2_type"
repeat_region 18703 .18828
/rpt_family="MER2_type"
repeat_region 19952 .20261
/rpt_family="L1"
repeat_region 20364 .20509
/rpt_family="Alu"
repeat_region 20841 .21461
/rpt_family="L1"
repeat_region 21879 .22025
/rpt_family="Alu"
repeat_region 22041 .22148
/rpt_family="MALR"
repeat_region 22149 .22501
/rpt_family="MALR"
repeat_region 22302 .22533
/rpt_family="MALR"
repeat_region 22712 .22988
/rpt_family="L1"
repeat_region 23969 .24065
/rpt_family="MALR"
repeat_region 24058 .24181
/rpt_family="MALR"
repeat_region 25028 .25216
/rpt_family="MALR"
repeat_region 26526 .26636
/rpt_family="MALR"
repeat_region 26682 .26919
/rpt_family="ERV1"
repeat_region 26973 .27576
/rpt_family="ERV1"
repeat_region 27986 .28021
/rpt_family="ERV1"
repeat_region 28028 .30206
/rpt_family="ERV1"
repeat_region 30666 .30758
/rpt_family="MALR"
repeat_region 30955 .31009
/rpt_family="MALR"
repeat_region 31260 .31397
/rpt_family="B4"
repeat_region 31572 .31710
/rpt_family="Alu"
```

```
repeat_region 32227 .32297
/rpt_family="B4"
repeat_region 32306 .32447
/rpt_family="Alu"
unare
repeat_region 35042 .35191
/note="Sequence derived from one plasmid subclone."
repeat_region 37307 .37672
/rpt_family="ERV1"
repeat_region 37938 .38097
/rpt_family="ERV1"
repeat_region 38359 .38608
/rpt_family="B4"
repeat_region 38875 .38986
```

Query Match 91.6%; Score 17.4; DB 10; Length 144026;  
Best Local Similarity 94.7%; Pred. No. 2.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGAGATGG 19  
Db 63490 TGGGGTTCGCTGAGATGG 63472

RESULT 15  
AC147366 175781 bp DNA linear ROD 13-JUL-2004  
LOCATION Mus musculus chromosome 15 clone RP23-386J3, complete sequence.  
DEFINITION AC147366  
VERSION AC147366.5 GI:50253696  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE 1 (bases 1 to 175781)  
JOURNAL Wilson, R.K.

REFERENCE  
AUTHORS The sequence of Mus musculus clone  
TITLE 2 (bases 1 to 175781)  
JOURNAL Wilson, R.K.

REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (11-NOV-2003) Genome Sequencing Center, 4444 Forest Park  
JOURNAL Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 175781)

REFERENCE  
AUTHORS Wilson, R.K.  
TITLE Direct Submission  
JOURNAL Submitted (25-MAY-2004) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 175781)

REFERENCE  
AUTHORS Wilson, R.K.  
TITLE Direct Submission  
JOURNAL Submitted (13-JUL-2004) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
On Jul 13, 2004 this sequence version replaced gi:47605149.

COMMENT  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submissions@wustl.edu  
----- Project Information -----  
Center project name: M\_BA0386J03  
-----

FEATURES  
source Location/Qualifiers  
1..175781

ORGANISM  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="15"  
/clone="RP23-386J3"

Query Match 91.6%; Score 17.4; DB 10; Length 175781;

Best Local Similarity 94.7%; Pred. No. 2.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGTTGGCTGGAGTGG 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 8125 TGGGTTCTCTGGAGTGG 8143

Search completed: November 7, 2005, 06:42:07  
Job time : 754.66 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21, Search time 219.106 Seconds  
(without alignments)  
513.336 Million cell updates/sec

Title: US-10-777-131A-4  
Perfect score: 19  
Sequence: 1 tggggctgcgcggaagtcg 19

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	19	ABZ79930	Abz79930 Mycobacte
2	19	100.0	1144	ABZ79953	Abz79953 Mycobacte
3	19	100.0	2406	ABZ79958	Abz79958 Mycobacte
4	19	100.0	2409	AAK34178	AAK34178 Mycobacte
5	19	100.0	2451	AAK34179	AAK34179 Mycobacte
6	19	100.0	2806	ABZ79957	Abz79957 Mycobacte
7	19	100.0	103765	AA199683_43	Continuation (44 o
8	19	100.0	110000	AA199682_43	Continuation (44 o
9	16.4	86.3	4428	ADN04195	Adn04195 Antispori
10	16.4	86.3	117328	ABD32886	Abd32886 Mouse can
11	15.8	83.2	292	AA125090	AA125090 Probe #15
12	15.8	83.2	292	ABA70776	AbA70776 Human foe
13	15.8	83.2	292	AA150949	AA150949 Probe #19
14	15.8	83.2	292	ABA37266	AbA37266 Probe #15
15	15.8	83.2	292	AAK44980	AAK44980 Human bon
16	15.8	83.2	292	AAK19029	AAK19029 Human bra
17	15.8	83.2	292	ABK44646	ABK44646 Human liv
18	15.8	83.2	292	ABK19225	ABK19225 Human gen
19	15.8	83.2	304	AA102296	AA102296 Human rep
20	15.8	83.2	445	AA190253	AA190253 Human pol

21	15.8	83.2	476	AA115873	AA115873 Probe #58
22	15.8	83.2	476	ABK58175	ABK58175 Human foe
23	15.8	83.2	476	AA137782	AA137782 Probe #54
24	15.8	83.2	476	ABA27360	ABA27360 Probe #58
25	15.8	83.2	476	AAK31915	AAK31915 Human bon
26	15.8	83.2	476	AAK06255	AAK06255 Human bra
27	15.8	83.2	476	ABK31605	ABK31605 Human liv
28	15.8	83.2	476	ABK06677	ABK06677 Human gen
29	15.8	83.2	738	ACB32274	ACB32274 Prokaryot
30	15.8	83.2	984	ADT42231	ADT42231 Bacteri
31	15.8	83.2	1814	AB199704	AB199704 Mouse isc
32	15.8	83.2	2213	ACC46652	ACC46652 Human dit
33	15.8	83.2	2511	ACA24500	ACA24500 Prokaryot
34	15.8	83.2	2725	ADB68895	ADB68895 C. neofo
35	15.8	83.2	3240	ADB62016	ADB62016 Human CDN
36	15.8	83.2	3268	AAK87591	AAK87591 DNA encod
37	15.8	83.2	5109	ACN42259	ACN42259 Human dia
38	15.8	83.2	5229	ACN42258	ACN42258 Human dia
39	15.8	83.2	5342	AAD00374	AAD00374 Rat smoot
40	15.8	83.2	5739	ABK70271	ABK70271 Human Jun
41	15.8	83.2	6327	AAK32301	AAK32301 Dermatomy
42	15.8	83.2	6417	ABL64410	ABL64410 Stomach C
43	15.8	83.2	6417	ABX04169	ABX04169 Human mRN
44	15.8	83.2	6417	ABK84696	ABK84696 Human CDN
45	15.8	83.2	6417	AAK54631	AAK54631 Human chr

## ALIGNMENTS

RESULT 1	
ID	ABZ79930 standard; DNA, 19 BP.
XX	ABZ79930;
AC	19-MAY-2003 (first entry)
XX	
DT	
XX	
DE	Mycobacterium tuberculosis Rv3908-2 PCR primer SEQ ID NO:4.
XX	
KW	Mycobacterium tuberculosis; mutT2; alka; ogt; Rv3908; mutY; Rv3909;
KM	detection; multidrug resistance; multiple drug resistance; MDR;
KX	infection; PCR primer; ss.
OS	Mycobacterium tuberculosis.
XX	
PN	WO2003016562-A2.
XX	
PD	27-FEB-2003.
XX	
PF	14-AUG-2002; 2002WO-EP009679.
XX	
PR	14-AUG-2001; 2001US-0311824P.
XX	
PR	21-AUG-2001; 2001US-0313523P.
XX	
PA	(INSP ) INST PASTEUR.
XX	
PI	Gicquel B;
XX	
DR	WPI; 2003-256711/25.
XX	
PT	Predicting the epidemic character of a Mycobacterium tuberculosis isolate
XX	and/or the acquisition of multiple drug resistance (MDR) by the isolate
PT	by detecting an alteration in the DNA repair system of the isolate.
XX	
PS	Claim 32; Page 16; 83pp; English.
XX	
CC	The present invention describes a method for predicting the epidemic
CC	character of a Mycobacterium tuberculosis isolate and/or a selective
CC	advantage to be maintained in the host and/or the acquisition of multiple
CC	drug resistance (MDR) by the isolate comprising detecting an alteration
CC	in the DNA repair system of the isolate. Also described: (1) detecting a

CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a PCR primer for M. tuberculosis RV3908,  
CC which is used in the exemplification of the present invention  
XX  
SQ Sequence 19 BP; 2 A; 2 C; 10 G; 5 T; 0 U; 0 Other;  
Query Match 100.0%; Score 19; DB 8; Length 19;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TGGGGTTCGCTGGAAGTGG 19  
DB 1 TGGGGTTCGCTGGAAGTGG 19  
RESULT 2  
ABZ79953/c 100.0%; Score 19; DB 8; Length 19;  
ID ABZ79953 standard; DNA; 1144 BP.  
XX  
AC ABZ79953;  
XX  
DT 19-MAY-2003 (first entry)  
XX  
DE Mycobacterium tuberculosis RV3908 nucleotide sequence SEQ ID NO:20.  
XX  
KM Mycobacterium tuberculosis; mutT2; alkA; ogt; RV3908; mutY; RV3909;  
KM detection; multidrug resistance; multiple drug resistance; MDR;  
KM infection; gene; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO2003016562-A2.  
XX  
PD 27-FEB-2003.  
XX  
PE 14-AUG-2002; 2002WO-EP009679.  
XX  
PR 14-AUG-2001; 2001US-0311824P.  
XX 21-AUG-2001; 2001US-0313523P.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Glcquel B;  
XX  
DR WPI; 2003-256711/25.  
XX  
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate  
PT by detecting an alteration in the DNA repair system of the isolate.  
XX  
PS Disclosure; Fig 5A; 83pp; English.  
XX  
CC The present invention describes a method for predicting the epidemic  
CC character of a Mycobacterium tuberculosis isolate and/or a selective  
CC advantage to be maintained in the host and/or the acquisition of multiple  
CC drug resistance (MDR) by the isolate comprising detecting an alteration  
CC in the DNA repair system of the isolate. Also described: (1) detecting a  
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) detecting a  
CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a M. tuberculosis RV3908 nucleotide

CC sequence, which is used in the exemplification of the present invention  
XX  
SQ Sequence 1144 BP; 194 A; 405 C; 379 G; 166 T; 0 U; 0 Other;  
Query Match 100.0%; Score 19; DB 8; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TGGGGTTCGCTGGAAGTGG 19  
DB 1143 TGGGGTTCGCTGGAAGTGG 1125  
RESULT 3  
ABZ79958/c 100.0%; Score 19; DB 8; Length 2406;  
ID ABZ79958 standard; DNA; 2406 BP.  
XX  
AC ABZ79958;  
XX  
DT 19-MAY-2003 (first entry)  
XX  
DE Mycobacterium tuberculosis RV3909 ORF sequence SEQ ID NO:32.  
XX  
KM Mycobacterium tuberculosis; mutT2; alkA; ogt; RV3908; mutY; RV3909;  
KM detection; multidrug resistance; multiple drug resistance; MDR;  
KM infection; gene; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO2003016562-A2.  
XX  
PD 27-FEB-2003.  
XX  
PE 14-AUG-2002; 2002WO-EP009679.  
XX  
PR 14-AUG-2001; 2001US-0311824P.  
XX 21-AUG-2001; 2001US-0313523P.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Glcquel B;  
XX  
DR WPI; 2003-256711/25.  
XX  
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate  
PT by detecting an alteration in the DNA repair system of the isolate.  
XX  
PS Claim 45; Fig 7C; 83pp; English.  
XX  
CC The present invention describes a method for predicting the epidemic  
CC character of a Mycobacterium tuberculosis isolate and/or a selective  
CC advantage to be maintained in the host and/or the acquisition of multiple  
CC drug resistance (MDR) by the isolate comprising detecting an alteration  
CC in the DNA repair system of the isolate. Also described: (1) detecting a  
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a M. tuberculosis RV3909 open reading  
CC frame (ORF) sequence, which is used in the exemplification of the present  
CC invention  
XX  
SQ Sequence 2406 BP; 369 A; 899 C; 760 G; 378 T; 0 U; 0 Other;  
Query Match 100.0%; Score 19; DB 8; Length 2406;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTCGCTGGAAGTGG 19  
|||||  
DB 200 TGGGGTTCGCTGGAAGTGG 182

## RESULT 4

AAK34178/c  
ID AAK34178 standard; DNA; 2409 BP.

AC AAK34178;

DT 06-JUL-1999 (first entry)

DE Mycobacterium species nucleic acid sequence 36D.

KM Secreted protein; Mycobacterium; primer; PCR; amplification; probe;  
hybridisation; detection; vaccine; immunisation; infection; ss.

OS Mycobacterium sp.

PN WO9909186-A2.

PD 25-FEB-1999.

PF 14-AUG-1998; 98WO-FR001813.

PR 14-AUG-1997; 97FR-00010404.

PR 11-SEP-1997; 97FR-00011325.

PA (INSP ) INST PASTEUR.

PI Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;  
Coguet De La Salmoniere Y;

DR WPI; 1999-181045/15.

DR P-PSDB; AAY04925.

PT Mycobacterial DNA vectors containing reporter constructs - for  
identifying coding or promoter sequences involved in infection-associated  
protein expression.

PS Claim 22; Fig 36D; 309pp; French.

CC Sequences AAK34001-X34252 represent nucleic acids encoding secreted  
proteins from various Mycobacterium species microorganisms. The  
nucleotide sequences can be used as primers and probes for methods for  
detecting and identifying mycobacteria, especially belonging to the M.  
tuberculosis complex. The encoded proteins can be used in vaccines for  
immunisation against a bacterial or viral infection

SQ Sequence 2409 BP; 370 A; 899 C; 761 G; 379 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 2; Length 2409;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTCGCTGGAAGTGG 19  
|||||  
DB 200 TGGGGTTCGCTGGAAGTGG 182

## RESULT 5

AAK34179/c  
ID AAK34179 standard; DNA; 2451 BP.

AC AAK34179;

DT 06-JUL-1999 (first entry)

DE Mycobacterium species nucleic acid sequence 36F.

KM Secreted protein; Mycobacterium; primer; PCR; amplification; probe;  
hybridisation; detection; vaccine; immunisation; infection; ss.

XX Mycobacterium sp.

PN WO9909186-A2.

PD 25-FEB-1999.

PF 14-AUG-1998; 98WO-FR001813.

PR 14-AUG-1997; 97FR-00010404.

PR 11-SEP-1997; 97FR-00011325.

PA (INSP ) INST PASTEUR.

PI Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;  
Coguet De La Salmoniere Y;

DR WPI; 1999-181045/15.

DR P-PSDB; AAY04926.

PT Mycobacterial DNA vectors containing reporter constructs - for  
identifying coding or promoter sequences involved in infection-associated  
protein expression.

PS Claim 22; Fig 36F; 309pp; French.

CC Sequences AAK34001-X34252 represent nucleic acids encoding secreted  
proteins from various Mycobacterium species microorganisms. The  
nucleotide sequences can be used as primers and probes for methods for  
detecting and identifying mycobacteria, especially belonging to the M.  
tuberculosis complex. The encoded proteins can be used in vaccines for  
immunisation against a bacterial or viral infection

SQ Sequence 2451 BP; 377 A; 915 C; 776 G; 383 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 2; Length 2451;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTCGCTGGAAGTGG 19  
|||||  
DB 242 TGGGGTTCGCTGGAAGTGG 224

## RESULT 6

ABZ79957/c  
ID ABZ79957 standard; DNA; 2806 BP.

AC ABZ79957;

DT 19-MAY-2003 (first entry)

DE Mycobacterium tuberculosis Rv3909 nucleotide sequence SEQ ID NO:22.

KM Mycobacterium tuberculosis; mutT2; alxA; ogt; Rv3908; mutY; Rv3909;  
infection; multidrug resistance; multiple drug resistance; MDR;

OS Mycobacterium tuberculosis.

PN WO2003016562-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-BP009679.

PR 14-AUG-2001; 2001US-0311824P.

PR 21-AUG-2001; 2001US-0313523P.

PA (INSP ) INST PASTEUR.

PI Gicquel B;

DR WP1; 2003-256711/25.  
XX Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate  
PT by detecting an alteration in the DNA repair system of the isolate.  
PS Disclosure; Fig 7A-B; 83pp; English.  
XX The present invention describes a method for predicting the epidemic  
CC character of a Mycobacterium tuberculosis isolate and/or a selective  
CC advantage to be maintained in the host and/or the acquisition of multiple  
CC drug resistance (MDR) by the isolate comprising detecting an alteration  
CC in the DNA repair system of the isolate. Also described: (1) detecting a  
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a M. tuberculosis Rv3909 nucleotide  
CC sequence, which is used in the exemplification of the present invention  
CC  
XX Sequence 2806 BP; 433 A; 1048 C; 886 G; 439 T; 0 U; 0 Other;  
SQ

Query Match 100.0%; Score 19; DB 8; Length 2806;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTG 19  
Db 400 TGGGGTTCGCTGGAAGTG 382

## RESULT 7

AI199683\_43/C  
Continuation (44 of 44) of AI199683 from base 4300001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AI199683 Accession AI199683

WP	Fragment Name	Begin	End
WP	AI199683_00	1	110000
WP	AI199683_01	100001	210000
WP	AI199683_02	200001	310000
WP	AI199683_03	300001	410000
WP	AI199683_04	400001	510000
WP	AI199683_05	500001	610000
WP	AI199683_06	600001	710000
WP	AI199683_07	700001	810000
WP	AI199683_08	800001	910000
WP	AI199683_09	900001	1010000
WP	AI199683_10	1000001	1110000
WP	AI199683_11	1100001	1210000
WP	AI199683_12	1200001	1310000
WP	AI199683_13	1300001	1410000
WP	AI199683_14	1400001	1510000
WP	AI199683_15	1500001	1610000
WP	AI199683_16	1600001	1710000
WP	AI199683_17	1700001	1810000
WP	AI199683_18	1800001	1910000
WP	AI199683_19	1900001	2010000
WP	AI199683_20	2000001	2110000
WP	AI199683_21	2100001	2210000
WP	AI199683_22	2200001	2310000
WP	AI199683_23	2300001	2410000
WP	AI199683_24	2400001	2510000
WP	AI199683_25	2500001	2610000
WP	AI199683_26	2600001	2710000
WP	AI199683_27	2700001	2810000
WP	AI199683_28	2800001	2910000
WP	AI199683_29	2900001	3010000
WP	AI199683_30	3000001	3110000
WP	AI199683_31	3100001	3210000
WP	AI199683_32	3200001	3310000

WP	AI199683_33	3300001	3410000
WP	AI199683_34	3400001	3510000
WP	AI199683_35	3500001	3610000
WP	AI199683_36	3600001	3710000
WP	AI199683_37	3700001	3810000
WP	AI199683_38	3800001	3910000
WP	AI199683_39	3900001	4010000
WP	AI199683_40	4000001	4110000
WP	AI199683_41	4100001	4210000
WP	AI199683_42	4200001	4310000
WP	AI199683_43	4300001	4403765

Query Match 100.0%; Score 19; DB 4; Length 103765;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTG 19  
Db 86625 TGGGGTTCGCTGGAAGTG 86607

## RESULT 8

AI199682\_43/C  
Continuation (44 of 45) of AI199682 from base 4300001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AI199682 Accession AI199682

WP	Fragment Name	Begin	End
WP	AI199682_00	1	110000
WP	AI199682_01	100001	210000
WP	AI199682_02	200001	310000
WP	AI199682_03	300001	410000
WP	AI199682_04	400001	510000
WP	AI199682_05	500001	610000
WP	AI199682_06	600001	710000
WP	AI199682_07	700001	810000
WP	AI199682_08	800001	910000
WP	AI199682_09	900001	1010000
WP	AI199682_10	1000001	1110000
WP	AI199682_11	1100001	1210000
WP	AI199682_12	1200001	1310000
WP	AI199682_13	1300001	1410000
WP	AI199682_14	1400001	1510000
WP	AI199682_15	1500001	1610000
WP	AI199682_16	1600001	1710000
WP	AI199682_17	1700001	1810000
WP	AI199682_18	1800001	1910000
WP	AI199682_19	1900001	2010000
WP	AI199682_20	2000001	2110000
WP	AI199682_21	2100001	2210000
WP	AI199682_22	2200001	2310000
WP	AI199682_23	2300001	2410000
WP	AI199682_24	2400001	2510000
WP	AI199682_25	2500001	2610000
WP	AI199682_26	2600001	2710000
WP	AI199682_27	2700001	2810000
WP	AI199682_28	2800001	2910000
WP	AI199682_29	2900001	3010000
WP	AI199682_30	3000001	3110000
WP	AI199682_31	3100001	3210000
WP	AI199682_32	3200001	3310000
WP	AI199682_33	3300001	3410000
WP	AI199682_34	3400001	3510000
WP	AI199682_35	3500001	3610000
WP	AI199682_36	3600001	3710000
WP	AI199682_37	3700001	3810000
WP	AI199682_38	3800001	3910000
WP	AI199682_39	3900001	4010000
WP	AI199682_40	4000001	4110000
WP	AI199682_41	4100001	4210000
WP	AI199682_42	4200001	4310000
WP	AI199682_43	4300001	4411529
WP	AI199682_44	4400001	

Query Match 100.0%; Score 19; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 19  
|||||  
DB 94388 TGGGGTTGCTGGAAGTGG 94370

## RESULT 9

ADN04199  
ID ADN04199 standard; cDNA; 4428 BP.

AC ADN04199;

DT 01-JUL-2004 (first entry)

DE Antipsoriatic cDNA sequence #299.

ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.

OS Homo sapiens.

PN WO2004028479-A2.

PD 08-APR-2004.

PF 25-SEP-2003; 2003WO-US030907.

PR 25-SEP-2002; 2002US-0414006P.

PA (GENTH ) GENENTECH INC.

PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
PI Wu TD;

DR WPI; 2004-305105/28.

DR P-PSDB; ADN04200.

PT New PRO nucleic acid or polypeptide, useful for preparing a  
PT pharmaceutical composition for diagnosing or treating psoriasis in a  
PT mammal.

PS Claim 1; SEQ ID NO 593; 3069pp; English.

CC The invention relates to novel polynucleotide and polypeptides for  
CC treating psoriasis or a sequence having at least 80% identity to the  
CC above sequences. The nucleic acid is useful for preparing a composition  
CC for diagnosing or treating psoriasis in a mammal. This sequence  
CC corresponds to one of the polynucleotides of the invention.

SQ Sequence 4428 BP; 1174 A; 1006 C; 1009 G; 1239 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 12; Length 4428;

Best Local Similarity 94.4%; Pred. No. 3e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 18  
|||||  
DB 753 TGGGGTTGCTGGAAGTGG 770

## RESULT 10

ABD32886/C  
ID ABD32886 standard; DNA; 117328 BP.

AC ABD32886;

DT 18-NOV-2004 (first entry)

DE Mouse cancer-associated genomic DNA MD18-013.

Mouse; ds; cancer-associated protein; gene; cytosolic; cancer;  
leukaemia; lymphoma; GAP.

XX Mus musculus.

OS WO2004074320-A2.

PN 02-SEP-2004.

PF 17-FEB-2004; 2004WO-US004730.

PR 14-FEB-2003; 2003US-00367094.

PR 15-APR-2003; 2003US-00388838.

PR 13-JUN-2003; 2003US-00417375.

PR 15-SEP-2003; 2003US-00661862.

PR 15-DEC-2003; 2003US-00737318.

PA (SAGR-) SAGRES DISCOVERY INC.

PI Morris DW, Morris DW, Malandro MS;

PI WPI; 2004-652914/63.

PT New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and  
PT leukemia, or in screening for agents that modulate cancer.

PS disclosure; seqid 577; 310pp; English.

CC The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acid encodes cancer-  
CC associated proteins. Also included are an expression vector comprising  
CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer  
CC or for detecting the presence or absence of cancer cells in an  
CC individual, a method for inhibiting growth of cancer cells in an  
CC individual, a method for delivering a therapeutic agent to cancer cells  
CC in an individual, an electronic library comprising the above  
CC polynucleotide or polypeptide (or their fragments), methods of screening  
CC for anticancer activity or for a bioactive agent capable of modulating  
CC the activity of a CA protein (GAP), methods for detecting cancer  
CC associated with expression of a polypeptide in a test cell sample, a  
CC method for treating cancers and a method for inhibiting the expression of  
CC CA gene in a cell. The composition and methods are useful for detecting,  
CC diagnosing, preventing and treating cancers, especially lymphoma and  
CC leukaemia. These may also be used in screening for agents that modulate  
CC cancer. The present sequence is a mouse GAP genomic sequence. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 117328 BP; 31499 A; 25773 C; 25903 G; 33390 T; 0 U; 763 Other;

Query Match 86.3%; Score 16.4; DB 13; Length 117328;

Best Local Similarity 94.4%; Pred. No. 3.6e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGTTGCTGGAAGTGG 19  
|||||  
DB 14283 GGGGTTGCTGGAAGTGG 14266

## RESULT 11

```

AAI25090
ID AAI25090 standard; DNA; 292 BP.
XX
AC AAI25090;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #15023 for gene expression analysis in human cervical cell sample.
XX
KM Probe; human; microarray; gene expression; cervical epithelial cell;
KM cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI, 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX
PS Claim 25; SEQ ID NO 15023; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 15.8; DB 4; Length 292;
Best Local Similarity 89.5%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TGGGGTTCGCTGGAAGTG 19
DB 30 TGGGATCTCTCGAAGTG 48
XX
RESULT 12
ABA70776
ID ABA70776 standard; DNA; 292 BP.
XX
AC ABA70776;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #19081.
XX
KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX

```

```

OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI, 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX
PS Claim 4; SEQ ID NO 19081; 639bp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 15.8; DB 4; Length 292;
Best Local Similarity 89.5%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TGGGGTTCGCTGGAAGTG 19
DB 30 TGGGATCTCTCGAAGTG 48
XX
RESULT 13
AAI50949
ID AAI50949 standard; DNA; 292 BP.
XX
AC AAI50949;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #19635 used to measure gene expression in human placenta sample.
XX
KM Probe; microarray; human; placenta; antenatal diagnosis;
KM genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
XX

```



PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488897/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.  
XX  
PS Claim 25; SEQ ID NO 19635; 654BP; English.  
XX  
CC The present invention relates to single exon nucleic acid probes (SENP).  
CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders  
XX  
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;  
XX  
Query Match 83.2%; Score 15.8; DB 4; Length 292;  
Best Local Similarity 89.5%; Pred. No. 5.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
QY 1 TGGGGTTTCGCTGGAAGTGG 19  
DB 30 TGGGGATCTCTGGAAGTGG 48  
XX  
RESULT 14  
ABA37266  
ID ABA37266 standard; DNA; 292 BP.  
XX  
AC ABA37266;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Probe #15732 for gene expression analysis in human heart cell sample.  
XX  
KW Human; gene expression; heart; microarray; vascular system; probe;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000666.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts.  
XX  
PS Claim 4; SEQ ID NO 15732; 530BP; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease. Note: The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at [http://wipo.int/pat/published\\_pct\\_sequences](http://wipo.int/pat/published_pct_sequences)  
XX  
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;  
XX  
Query Match 83.2%; Score 15.8; DB 4; Length 292;  
Best Local Similarity 89.5%; Pred. No. 5.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
QY 1 TGGGGTTTCGCTGGAAGTGG 19  
DB 30 TGGGGATCTCTGGAAGTGG 48  
XX  
RESULT 15  
AAK4980  
ID AAK4980 standard; DNA; 292 BP.  
XX  
AC AAK4980;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 19537.  
XX  
DE Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000668.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488900/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.  
XX  
PS Example 4; SEQ ID NO 19537; 658BP + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention  
XX  
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;  
XX  
Query Match 83.2%; Score 15.8; DB 4; Length 292;  
Best Local Similarity 89.5%; Pred. No. 5.1e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TGGGTTGCTGGAGTGG 19  
 |||||  
 Db 30 TGGGATCTTGGAGTGG 48  
 |||||

Search completed: November 7, 2005, 05:40:33  
 Job time : 233.106 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: November 6, 2005, 19:19:37 ; Search time 64.4787 Seconds  
(without alignments)  
482.163 Million cell updates/sec

Title: US-10-777-131a-4

Sequence: 1 tgggggttcgctggaagtgg 19

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A COMB. seq: \*  
2: /cgn2\_6/prodata/1/ina/5B COMB. seq: \*  
3: /cgn2\_6/prodata/1/ina/6A COMB. seq: \*  
4: /cgn2\_6/prodata/1/ina/6B COMB. seq: \*  
5: /cgn2\_6/prodata/1/ina/PCTUS COMB. seq: \*  
6: /cgn2\_6/prodata/1/ina/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 2	19	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 3	15.8	83.2	2678	4 US-09-807-757C-8	Sequence 8, Appli
C 4	15.8	83.2	5342	4 US-09-807-757C-1	Sequence 1, Appli
C 5	15.8	83.2	6328	3 US-08-913-832A-1	Sequence 1, Appli
C 6	15.8	83.2	6328	3 US-09-249-181A-1	Sequence 1, Appli
C 7	15.8	83.2	6328	4 US-09-158-707-1	Sequence 325, Appli
C 8	15.8	83.2	6475	4 US-09-620-312D-325	Sequence 13910, A
C 9	15.8	83.2	40897	4 US-09-949-016-13910	Sequence 13911, A
C 10	15.8	83.2	40897	4 US-09-949-016-13911	Sequence 13912, A
C 11	15.8	83.2	40897	4 US-09-949-016-13912	Sequence 13913, A
C 12	15.8	83.2	40897	4 US-09-949-016-13913	Sequence 14409, A
C 13	15.8	83.2	40897	4 US-09-949-016-14409	Sequence 14410, A
C 14	15.8	83.2	40897	4 US-09-949-016-14410	Sequence 14411, A
C 15	15.8	83.2	40897	4 US-09-949-016-14411	Sequence 14412, A
C 16	15.8	83.2	40897	4 US-09-949-016-14412	Sequence 15578, A
C 17	15.8	83.2	162914	4 US-09-949-016-15578	Sequence 15579, A
C 18	15.4	81.1	570	4 US-09-621-976-313	Sequence 313, Appli
C 19	15.4	81.1	872	4 US-09-786-240-23	Sequence 23, Appli
C 20	15.4	81.1	1007	4 US-09-949-016-917	Sequence 917, Appli
C 21	15.4	81.1	1007	4 US-09-949-016-3973	Sequence 3973, Appli
C 22	15.4	81.1	4773	4 US-09-949-016-2552	Sequence 2552, Appli
C 23	15.4	81.1	6415	4 US-09-949-016-1015	Sequence 1015, Appli
C 24	15.4	81.1	7388	4 US-09-949-016-15801	Sequence 15901, A
C 25	15.4	81.1	80858	4 US-09-949-016-12659	Sequence 12659, A
C 26	15.4	81.1	80859	4 US-09-949-016-15715	Sequence 15715, A
C 27	15	78.9	965	4 US-09-270-767-9078	Sequence 9078, Appli

C 28	15	78.9	965	4 US-09-270-767-24360	Sequence 24360, A
C 29	14.8	77.9	567	4 US-09-893-737-15	Sequence 15, Appli
C 30	14.8	77.9	601	4 US-09-949-016-71607	Sequence 71607, A
C 31	14.8	77.9	786	4 US-09-902-540-3126	Sequence 3126, Appli
C 32	14.8	77.9	817	4 US-09-809-545A-7	Sequence 7, Appli
C 33	14.8	77.9	1443	4 US-09-252-991A-1355	Sequence 3155, Appli
C 34	14.8	77.9	1929	4 US-09-252-991A-2815	Sequence 2815, Appli
C 35	14.8	77.9	9775	3 US-08-977-171-1	Sequence 1, Appli
C 36	14.8	77.9	15268	4 US-09-902-540-1142	Sequence 1142, Appli
C 37	14.8	77.9	31300	4 US-09-949-016-16367	Sequence 16367, A
C 38	14.8	77.9	32065	4 US-09-949-016-12136	Sequence 12136, A
C 39	14.8	77.9	32066	4 US-09-949-016-12136	Sequence 13268, A
C 40	14.8	77.9	38371	4 US-09-949-016-12061	Sequence 12061, A
C 41	14.8	77.9	38371	4 US-09-949-016-12488	Sequence 12488, A
C 42	14.8	77.9	38371	4 US-09-949-016-15596	Sequence 15596, A
C 43	14.8	77.9	38371	4 US-09-949-016-15597	Sequence 15597, A
C 44	14.8	77.9	46885	4 US-09-949-016-13848	Sequence 13848, A
C 45	14.8	77.9	50797	4 US-09-949-016-16346	Sequence 16346, A

#### ALIGNMENTS

```
RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match          100.0%; Score 19; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGGGTTCGCTGGAAGTGG 19
Db      4386625 TGGGGTTCGCTGGAAGTGG 4386607

RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          100.0%; Score 19; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
DB      4394388 TGGGGTTGCTGGAAGTGG 4394370

RESULT 3
US-09-807-757C-8/c
; Sequence 8, Application US/09807757C
; Patent No. 6825035
; GENERAL INFORMATION:
; APPLICANT: Owens, Gary K.
; APPLICANT: Mack, Christopher
; APPLICANT: Blank, Randall
; APPLICANT: University of Virginia Patent Foundation
; TITLE OF INVENTION: Compositions and Methods for Modulating Expression
; FILE REFERENCE: 021258-000500US
; CURRENT APPLICATION NUMBER: US/09/807,757C
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 60/105,330
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: WO PCT/US99/24972
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2678
; TYPE: DNA
; ORGANISM: Rattus sp.
; FEATURE:
; OTHER INFORMATION: smooth muscle alpha-actin (SM alpha-A) gene
; OTHER INFORMATION: first intron sequence
US-09-807-757C-8

Query Match          83.2%; Score 15.8; DB 4; Length 2678;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
DB      1846 TGGGGTTCTTGGAGAGTGG 1828

RESULT 4
US-09-807-757C-1/c
; Sequence 1, Application US/09807757C
; Patent No. 6825035
; GENERAL INFORMATION:
; APPLICANT: Owens, Gary K.
; APPLICANT: Mack, Christopher
; APPLICANT: Blank, Randall
; APPLICANT: University of Virginia Patent Foundation
; TITLE OF INVENTION: Compositions and Methods for Modulating Expression
; FILE REFERENCE: 021258-000500US
; CURRENT APPLICATION NUMBER: US/09/807,757C
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 60/105,330
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: WO PCT/US99/24972
; PRIOR FILING DATE: 1999-10-22
```

```

; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 5342
; TYPE: DNA
; ORGANISM: Rattus sp.
; FEATURE:
; OTHER INFORMATION: smooth muscle alpha-actin (SM alpha-A) gene
; OTHER INFORMATION: regulatory region 5' promoter and intron
US-09-807-757C-1

Query Match          83.2%; Score 15.8; DB 4; Length 5342;
Best Local Similarity 89.5%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
DB      4452 TGGGGTTCTTGGAGAGTGG 4434

RESULT 5
US-08-913-832A-1/c
; Sequence 1, Application US/08913832A
; Patent No. 6329517
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0030-999
; CURRENT APPLICATION NUMBER: US/08/913,832A
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: PCT/DE96/00444
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6328
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(5736)
US-08-913-832A-1

Query Match          83.2%; Score 15.8; DB 3; Length 6328;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
DB      1453 TGGGATCTCTGGAAGTGG 1435

RESULT 6
US-09-249-181A-1/c
; Sequence 1, Application US/09249181A
; Patent No. 6440679
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0059-999
; CURRENT APPLICATION NUMBER: US/09/249,181A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/913,832
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: PCT/DE96/00444
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6328
```

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)... (5736)  
US-09-249-181A-1

Query Match  
Best Local Similarity 83.2%; Score 15.8; DB 3; Length 6328;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19  
Db 1453 TGGGGATCTCTGGAAGTGG 1435

RESULT 7  
US-09-158-707-1/c  
Sequence 1, Application US/09158707  
Patent No. 6500923

GENERAL INFORMATION:  
APPLICANT: Seelig, Hans Peter  
APPLICANT: Renz, Manfred  
TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN  
FILE REFERENCE: 8484-0043-999  
CURRENT APPLICATION NUMBER: US/09/158,707  
CURRENT FILING DATE: 1998-09-22  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 6328  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)... (5736)  
US-09-158-707-1

Query Match  
Best Local Similarity 83.2%; Score 15.8; DB 4; Length 6328;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19  
Db 1453 TGGGGATCTCTGGAAGTGG 1435

RESULT 8  
US-09-620-312D-325/c  
Sequence 325, Application US/09620312D  
Patent No. 6569662

GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyun  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonhong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yungling  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: John Tillinghaast  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/09/620,312D

CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: PL FL\_genes Version 1.0  
SEQ ID NO 325  
LENGTH: 6475  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (219)... (5957)  
US-09-620-312D-325

Query Match  
Best Local Similarity 83.2%; Score 15.8; DB 4; Length 6475;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19  
Db 1671 TGGGGATCTCTGGAAGTGG 1653

RESULT 9  
US-09-949-016-13910  
Sequence 13910, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13910  
LENGTH: 40897  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-13910

Query Match  
Best Local Similarity 83.2%; Score 15.8; DB 4; Length 40897;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19  
Db 1452 TGGGGATCTCTGGAAGTGG 1470

RESULT 10  
US-09-949-016-13911  
Sequence 13911, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768

```

1 PRIOR FILING DATE: 2000-10-03
2 PRIOR APPLICATION NUMBER: 60/231,498
3 PRIOR FILING DATE: 2000-09-08
4 NUMBER OF SEQ ID NOS: 207012
5 SOFTWARE: FASTSEQ for Windows Version 4.0
6 SEQ ID NO 13911
7 LENGTH: 40897
8 TYPE: DNA
9 ORGANISM: Human
10 OS-09-949-016-13911

```

Query Match	83.2%	Score 15.8;	DB 4;	length 40897;
Best Local Similarity	89.5%	Pred. No. 1.8e+02;		
Matches 17; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY 1 TGGGCTTCGCTGGAAGTGG 19  
||| ||||| |||||  
Db 1452 TGGCGGTCGCTGGAAGTGG 1470

```

RESULT 11
US-09-949-016-13912
/ Sequence 13912, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13912
/ LENGTH: 40857
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-13912

```

Query Match	83.2%	Score 15.8;	DB 4;	Length 40897;
Best Local Similarity	89.5%;	Pred. No. 1.8e+02;		
Matches 17; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

QY 1 TGGGGTTCGCTCGAAGTGC 19  
 |||||  
 Db 1452 TGGCGGTCGCTGGAAGTGC 1470

```

RESULT 12
US-09-949-016-13913
; Sequence 13913, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 13913
; LENGTH: 40897
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13913

```

Query Match	83.2%	Score 15.8;	DB 4;	Length 40897;
Best Local Similarity	89.5%	Pred. No. 1.8e+02;		
Matches 17; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY 1 TGGGTTCCCTGGAAGTGG 19  
Db 1452 TGGCGTCCGCTGGAAGTGG 1470

```

RESULT 13
US-09-949-016-14409
; Sequence 14409, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949.016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14409
; LENGTH: 40897
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14409

```

Query Match	83.2%	Score 15.8;	DB 4;	Length 40897;
Best Local Similarity	89.5%;	Pred. No. 1.8e+02;		
Matches 17;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0

QY 1 TGGGGTTCGCTGGAAGTGC 19  
||| | ||||| |||||  
Db 1452 TGGCGGTGCTGGAGTGC 1470

```

1      RESULT 14
2      US-09-949-016-14410
3      /
4      ; Sequence 14410, Application US/09949016
5      ; Patent No. 6812339
6      ; GENERAL INFORMATION:
7      ; APPLICANT: VENTER, J. Craig et al.
8      ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
9      ; TITLE OF INVENTION: WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF
10     ; FILE REFERENCE: CLO01307
11     ; CURRENT APPLICATION NUMBER: US/09/949, 016
12     ; CURRENT FILING DATE: 2000-04-14
13     ; PRIOR APPLICATION NUMBER: 60/241,755
14     ; PRIOR FILING DATE: 2000-10-20
15     ; PRIOR APPLICATION NUMBER: 60/237,768
16     ; PRIOR FILING DATE: 2000-10-03
17     ; PRIOR APPLICATION NUMBER: 60/231,498
18     ; PRIOR FILING DATE: 2000-03-08
19     ; NUMBER OF SEQ ID NOS: 207012
20     ; SOFTWARE: FastSeq for Windows Version 4.0
21     ; SEQ ID NO 14410
22     ; LENGTH: 40897
23     ;
24     ; TYPE: DNA
25     ; ORGANISM: Human
26     ; US-09-949-016-14410

```



This page blank (uspiq)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: November 6, 2005, 20:56:42 ; Search time 473.787 Seconds  
(without alignments)  
331.640 Million cell updates/sec

Title: US-10-777-131A-4

Perfect score: 19  
Sequence: 1 tggggctcgcgtgaagttg 19

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB2.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:\*  
23: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
24: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB2.seq:\*  
25: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq:\*  
26: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
27: /cgn2\_6/ptodata/2/pubpna/US66\_NEW\_PUB.seq:\*  
28: /cgn2\_6/ptodata/2/pubpna/US66\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	US-10-216-817-4	Sequence 4, Appli
2	19	100.0	19	US-10-777-131A-4	Sequence 4, Appli
3	19	100.0	1144	US-10-216-817-20	Sequence 20, Appli
4	19	100.0	1144	US-10-777-131A-20	Sequence 20, Appli
5	19	100.0	2406	US-10-216-817-32	Sequence 32, Appli

Result No.	Score	Query Match	Length	ID	Description
6	19	100.0	2406	US-10-777-131A-32	Sequence 32, Appli
7	19	100.0	2409	US-09-855-604-618	Sequence 618, App
8	19	100.0	2409	US-09-855-604-618	Sequence 618, App
9	19	100.0	2451	US-09-855-604-620	Sequence 620, App
10	19	100.0	2451	US-09-855-604-620	Sequence 620, App
11	19	100.0	2806	US-10-216-817-22	Sequence 22, Appli
12	19	100.0	2806	US-10-216-817-22	Sequence 22, Appli
13	19	100.0	105077	US-10-777-131A-22	Sequence 22, Appli
14	16.4	86.3	117328	US-10-417-375-15	Sequence 15, Appli
15	16	84.2	694	US-10-461-862-127	Sequence 127, App
16	16	84.2	694	US-10-027-632-150246	Sequence 150246,
17	16	84.2	1343	US-10-027-632-150246	Sequence 150246,
18	15.8	83.2	201	US-09-925-065A-670362	Sequence 670362,
19	15.8	83.2	201	US-10-741-600-51483	Sequence 51483, A
20	15.8	83.2	255	US-10-741-600-51485	Sequence 51485, A
21	15.8	83.2	292	US-10-424-559-42574	Sequence 42574, A
22	15.8	83.2	304	US-09-864-761-22586	Sequence 22586, A
23	15.8	83.2	319	US-09-764-891-2297	Sequence 2297, Ap
24	15.8	83.2	403	US-10-425-115-40863	Sequence 40863, A
25	15.8	83.2	435	US-10-425-115-110341	Sequence 110341,
26	15.8	83.2	435	US-10-027-632-96855	Sequence 96855, A
27	15.8	83.2	476	US-10-027-632-96855	Sequence 96855, A
28	15.8	83.2	584	US-09-864-761-5826	Sequence 5826, Ap
29	15.8	83.2	584	US-09-925-065A-45152	Sequence 45152, A
30	15.8	83.2	628	US-09-925-065A-45153	Sequence 45153, A
31	15.8	83.2	636	US-10-767-701-8112	Sequence 8112, Ap
32	15.8	83.2	636	US-09-925-065A-490314	Sequence 490314,
33	15.8	83.2	738	US-09-925-065A-490315	Sequence 490315,
34	15.8	83.2	984	US-10-282-122A-20144	Sequence 20144, A
35	15.8	83.2	1323	US-10-369-493-40669	Sequence 40669, A
36	15.8	83.2	1323	US-09-925-065A-682717	Sequence 682717,
37	15.8	83.2	2511	US-09-925-065A-682718	Sequence 682718,
38	15.8	83.2	2725	US-10-287-122A-12370	Sequence 12370, A
39	15.8	83.2	3240	US-10-320-797-22	Sequence 22, Appli
40	15.8	83.2	3268	US-10-104-047-170	Sequence 170, App
41	15.8	83.2	5739	US-10-450-763-23395	Sequence 23395, A
42	15.8	83.2	6417	US-09-960-255-142	Sequence 142, App
43	15.8	83.2	6417	US-09-962-436-288	Sequence 288, App
44	15.8	83.2	6417	US-10-096-534-28	Sequence 28, Appli
45	15.8	83.2	6417	US-10-843-611A-2747	Sequence 2747, App
				US-10-479-874A-6	Sequence 6, Appli

#### ALIGNMENTS

RESULT 1  
US-10-216-817-4  
; Sequence 4, Application US/10216817  
; Publication No. US20030129619A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; FILE REFERENCE: 03495.0233-00000  
; CURRENT APPLICATION NUMBER: US/10/216,817  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 19  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Primer  
US-10-216-817-4  
Query Match 100.0%; Score 19; DB 16; Length 19;  
Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19  
Db 1 TGGGGTTGCTGGAAGTGG 19

## RESULT 2

US-10-777-131A-4  
; Sequence 4, Application US/10777131A  
; Publication No. US20050026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICOUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT APPLICATION NUMBER: US/10/777,131A  
; PRIOR FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 4  
; LENGTH: 19  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer  
US-10-777-131A-4

Query Match 100.0%; Score 19; DB 22; Length 19;  
Best Local Similarity 100.0%; Pred. No. 8.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19  
Db 1 TGGGGTTGCTGGAAGTGG 19

RESULT 3  
US-10-216-817-20/c  
; Sequence 20, Application US/10216817  
; Publication No. US20030129619A1  
; GENERAL INFORMATION:  
; APPLICANT: GICOUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; FILE REFERENCE: 03495.0233-00000  
; CURRENT APPLICATION NUMBER: US/10/216,817  
; PRIOR FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 1144  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-216-817-20

Query Match 100.0%; Score 19; DB 16; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 7.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19

Db 1143 TGGGGTTGCTGGAAGTGG 1125

RESULT 4  
US-10-777-131A-20/c  
; Sequence 20, Application US/10777131A  
; Publication No. US20050026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICOUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT APPLICATION NUMBER: US/10/777,131A  
; PRIOR FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 20  
; LENGTH: 1144  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-777-131A-20

Query Match 100.0%; Score 19; DB 22; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 7.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19  
Db 1143 TGGGGTTGCTGGAAGTGG 1125

RESULT 5  
US-10-216-817-32/c  
; Sequence 32, Application US/10216817  
; Publication No. US20030129619A1  
; GENERAL INFORMATION:  
; APPLICANT: GICOUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; FILE REFERENCE: 03495.0233-00000  
; CURRENT APPLICATION NUMBER: US/10/216,817  
; PRIOR FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 2406  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-216-817-32

Query Match 100.0%; Score 19; DB 16; Length 2406;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19  
Db 200 TGGGGTTGCTGGAAGTGG 182

RESULT 6

US-10-777-131a-32/c  
; Sequence 32, Application US/10777131A  
; Publication No. US2005026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT APPLICATION NUMBER: US/10/777, 131A  
; CURRENT FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 32  
; LENGTH: 2406  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-777-131a-32

Query Match 100.0%; Score 19; DB 22; Length 2406;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 19  
|||||  
DB 200 TGGGGTTGCTGGAAGTGG 182

RESULT 7  
US-09-855-604-618/c  
; Sequence 618, Application US/09855604  
; Publication No. US20040214165A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: PORTNOI, DENIS  
; APPLICANT: LIM, ENG-MONG  
; APPLICANT: PELICIC, VLADIMIR  
; APPLICANT: GUIGUENO, AGNES  
; APPLICANT: GOGUET DE LA SALMONIERE, YVES  
; TITLE OF INVENTION: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA,  
; TITLE OF INVENTION: VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND  
; FILE REFERENCE: 03715.0062-01000  
; CURRENT APPLICATION NUMBER: US/09/855,604  
; CURRENT FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: 09/485,536  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: PCT/FR98/01813  
; PRIOR FILING DATE: 1998-08-14  
; PRIOR APPLICATION NUMBER: FR 97 10404  
; PRIOR FILING DATE: 1997-08-14  
; PRIOR APPLICATION NUMBER: FR 97 11325  
; PRIOR FILING DATE: 1997-09-11  
; NUMBER OF SEQ ID NOS: 935  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 618  
; LENGTH: 2409  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2406)  
US-09-855-604-618

Query Match 100.0%; Score 19; DB 11; Length 2409;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 19  
|||||  
DB 200 TGGGGTTGCTGGAAGTGG 182

RESULT 8  
US-09-855-604-618/c  
; Sequence 618, Application US/09855604  
; Publication No. US20050158714A3  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: PORTNOI, DENIS  
; APPLICANT: LIM, ENG-MONG  
; APPLICANT: PELICIC, VLADIMIR  
; APPLICANT: GUIGUENO, AGNES  
; APPLICANT: GOGUET DE LA SALMONIERE, YVES  
; TITLE OF INVENTION: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA,  
; TITLE OF INVENTION: VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND  
; FILE REFERENCE: 03715.0062-01000  
; CURRENT APPLICATION NUMBER: US/09/855,604  
; CURRENT FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: 09/485,536  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: PCT/FR98/01813  
; PRIOR FILING DATE: 1998-08-14  
; PRIOR APPLICATION NUMBER: FR 97 10404  
; PRIOR FILING DATE: 1997-08-14  
; PRIOR APPLICATION NUMBER: FR 97 11325  
; PRIOR FILING DATE: 1997-09-11  
; NUMBER OF SEQ ID NOS: 935  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 618  
; LENGTH: 2409  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2406)  
US-09-855-604-618

Query Match 100.0%; Score 19; DB 13; Length 2409;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 19  
|||||  
DB 200 TGGGGTTGCTGGAAGTGG 182

RESULT 9  
US-09-855-604-620/c  
; Sequence 620, Application US/09855604  
; Publication No. US20040214165A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: PORTNOI, DENIS  
; APPLICANT: LIM, ENG-MONG  
; APPLICANT: PELICIC, VLADIMIR  
; APPLICANT: GUIGUENO, AGNES  
; APPLICANT: GOGUET DE LA SALMONIERE, YVES  
; TITLE OF INVENTION: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA,  
; TITLE OF INVENTION: VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND  
; FILE REFERENCE: 03715.0062-01000  
; CURRENT APPLICATION NUMBER: US/09/855,604  
; CURRENT FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: 09/485,536  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: PCT/FR98/01813  
; PRIOR FILING DATE: 1998-08-14  
; PRIOR APPLICATION NUMBER: FR 97 10404

```
/ PRIOR FILING DATE: 1997-08-14
/ PRIOR APPLICATION NUMBER: FR 97 11325
/ PRIOR FILING DATE: 1997-09-11
/ NUMBER OF SEQ ID NOS: 935
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO: 620
/ LENGTH: 2451
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (4)..(2448)
US-09-855-604-620
```

```
Query Match          100.0%; Score 19; DB 11; Length 2451;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGGGTTTCGCTGGAAGTGG 19
         |||||
Db      242 TGGGGTTTCGCTGGAAGTGG 224
```

```
RESULT 10
US-09-855-604-620/c
/ Sequence 620, Application US/09855604
/ Publication No. US20050158714A9
/ GENERAL INFORMATION:
```

```
/ APPLICANT: GICQUEL, BRIGITTE
/ APPLICANT: PORTNOI, DENIS
/ APPLICANT: LIM, ENG-MONG
/ APPLICANT: PELICIC, VLADIMIR
/ APPLICANT: GUIGUENO, AGNES
/ APPLICANT: GOGUET DE LA SALMONIERE, YVES
/ TITLE OF INVENTION: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA,
/ TITLE OF INVENTION: VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND
/ TITLE OF INVENTION: PREVENTING TUBERCULOSIS
/ FILE REFERENCE: 03715.0062-01000
/ CURRENT APPLICATION NUMBER: US/09/855,604
/ CURRENT FILING DATE: 2001-05-16
/ PRIOR APPLICATION NUMBER: 09/485,536
/ PRIOR FILING DATE: 2000-02-14
/ PRIOR APPLICATION NUMBER: PCT/FR98/01813
/ PRIOR FILING DATE: 1998-08-14
/ PRIOR APPLICATION NUMBER: FR 97 10404
/ PRIOR FILING DATE: 1997-08-14
/ PRIOR APPLICATION NUMBER: FR 97 11325
/ PRIOR FILING DATE: 1997-09-11
/ NUMBER OF SEQ ID NOS: 935
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO: 620
/ LENGTH: 2451
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (4)..(2448)
US-09-855-604-620
```

```
Query Match          100.0%; Score 19; DB 13; Length 2451;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGGGTTTCGCTGGAAGTGG 19
         |||||
Db      242 TGGGGTTTCGCTGGAAGTGG 224
```

```
RESULT 11
US-10-216-817-22/c
/ Sequence 22, Application US/10216817
/ Publication No. US20030129619A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: GICQUEL, BRIGITTE
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
/ TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
/ TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
/ FILE REFERENCE: 03495.0233-00000
/ CURRENT APPLICATION NUMBER: US/10/216,817
/ CURRENT FILING DATE: 2002-11-13
/ PRIOR APPLICATION NUMBER: 60/311,824
/ PRIOR FILING DATE: 2001-08-14
/ PRIOR APPLICATION NUMBER: 60/313,523
/ PRIOR FILING DATE: 2001-08-21
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO: 22
/ LENGTH: 2806
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
US-10-216-817-22
```

```
Query Match          100.0%; Score 19; DB 16; Length 2806;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGGGTTTCGCTGGAAGTGG 19
         |||||
Db      400 TGGGGTTTCGCTGGAAGTGG 382
```

```
RESULT 12
US-10-777-131a-22/c
/ Sequence 22, Application US/10777131A
/ Publication No. US20050026216A1
/ GENERAL INFORMATION:
/ APPLICANT: GICQUEL, BRIGITTE
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
/ TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
/ TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
/ FILE REFERENCE: 02356.0090-00000
/ CURRENT APPLICATION NUMBER: US/10/777,131A
/ CURRENT FILING DATE: 2004-02-13
/ PRIOR APPLICATION NUMBER: PCT/EP02/09679
/ PRIOR FILING DATE: 2002-08-14
/ PRIOR APPLICATION NUMBER: 60/311,824
/ PRIOR FILING DATE: 2001-08-14
/ PRIOR APPLICATION NUMBER: 60/313,523
/ PRIOR FILING DATE: 2001-08-21
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: PatentIn Ver. 3.2
/ SEQ ID NO: 22
/ LENGTH: 2806
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
US-10-777-131a-22
```

```
Query Match          100.0%; Score 19; DB 22; Length 2806;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGGGTTTCGCTGGAAGTGG 19
         |||||
Db      400 TGGGGTTTCGCTGGAAGTGG 382
```

```
RESULT 13
US-10-417-375-15/c
/ Sequence 15, Application US/10417375
/ Publication No. US20040219528A1
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc Malandro
/ TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
/ FILE REFERENCE: 529452001600
/ CURRENT APPLICATION NUMBER: US/10/417,375
```

```
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 105077
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(105077)
; OTHER INFORMATION: n = A,T,C or G
US-10-417-375-15
```

```
Query Match          91.6%; Score 17.4; DB 21; Length 105077;
Best Local Similarity 94.7%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 TGGGTTGCTGCTGGAAGTGG 19
DB      34989 TGGGTTACCTGCTGGAAGTGG 34971
```

```
RESULT 14
US-10-461-862-127/C
; Sequence 127, Application US/10461862
; Publication No. US20050904341
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001800
; CURRENT APPLICATION NUMBER: US/10/461,862
; CURRENT FILING DATE: 2003-06-13
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127
; LENGTH: 117328
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(117328)
; OTHER INFORMATION: n = A,T,C or G
US-10-461-862-127
```

```
Query Match          86.3%; Score 16.4; DB 22; Length 117328;
Best Local Similarity 94.4%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 GGGGTTGCTGCTGGAAGTGG 19
DB      14283 GGGGTTGCTGCTGGAAGTGG 14266
```

```
RESULT 15
US-10-027-632-150246
; Sequence 150246, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
```

```
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150246
; LENGTH: 694
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-150246
```

```
Query Match          84.2%; Score 16; DB 14; Length 694;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      4 GGTTCGCTGGAAGTGG 19
DB      199 GGTTCGCTGGAAGTGG 214
```

Search completed: November 7, 2005, 05:21:14  
Job time : 476.787 secs

*This Page Blank (uspi0)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 1900.81 Seconds  
(without alignments)  
380.481 Million cell updates/sec

Title: US-10-777-131a-4

Perfect score: 19  
Sequence: 1 tggggcttcgctggaagtg 19

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: gb\_esc1:\*  
2: gb\_esc2:\*  
3: gb\_hic:\*  
4: gb\_esc3:\*  
5: gb\_esc4:\*  
6: gb\_esc5:\*  
7: gb\_esc6:\*  
8: gb\_gsa1:\*  
9: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.4	91.6	495	5	BP055594 BP055594
2	17.4	91.6	508	1	AV778474 BP055594
3	17.4	91.6	512	5	BP028473 BP028473
4	17.4	91.6	519	1	AV779844 AV779844
5	17.4	91.6	523	1	AV779524 AV779524
6	17.4	91.6	526	4	BI419678 BI419678
7	17.4	91.6	528	1	AV779426 AV779426
8	17.4	91.6	531	1	AV778725 AV778725
9	17.4	91.6	532	1	BP043328 BP043328
10	17.4	91.6	536	1	AV777359 AV777359
11	17.4	91.6	552	1	AV777186 AV777186
12	17.4	91.6	557	5	BP055162 BP055162
13	17.4	91.6	565	5	BP055337 BP055337
14	17.4	91.6	565	5	BP056252 BP056252
15	17.4	91.6	583	5	BP034064 BP034064
16	17.4	91.6	604	1	AV776877 AV776877
17	17.4	91.6	607	1	AV777307 AV777307
18	17.4	91.6	638	8	AZ405854 AZ405854
19	17.4	91.6	785	5	BK626838 BK626838
20	17.4	91.6	915	4	BG178285 BG178285
21	17.4	91.6	960	9	CNS050735 CNS050735
22	16.4	86.3	354	6	BY550735 BY550735
23	16.4	86.3	413	5	BY310705 BY310705
24	16.4	86.3	490	5	BP057779 BP057779

C 25	16.4	86.3	510	5	BP052554 BP052554
C 26	16.4	86.3	517	5	BP053880 BP053880
C 27	16.4	86.3	526	5	BP065934 BP065934
C 28	16.4	86.3	706	1	AU089324 AU089324
C 29	16.4	86.3	742	2	BF143421 BF143421
C 30	16.4	86.3	922	9	CL507863 CL507863
C 31	16.4	86.3	1033	5	BQ925691 BQ925691
C 32	16.4	86.3	236	6	CB877779 CB877779
C 33	16.4	86.3	398	5	BQ759027 BQ759027
C 34	16.4	86.3	461	1	AV918034 AV918034
C 35	16.4	86.3	461	1	AV922614 AV922614
C 36	16.4	86.3	461	5	BQ660582 BQ660582
C 37	16.4	86.3	462	5	BQ664571 BQ664571
C 38	16.4	86.3	484	6	CA016770 CA016770
C 39	16.4	86.3	496	1	AV917430 AV917430
C 40	16.4	86.3	505	1	AV916508 AV916508
C 41	16.4	86.3	521	1	AV944743 AV944743
C 42	16.4	86.3	560	9	CL245534 CL245534
C 43	16.4	86.3	593	6	CB881023 CB881023
C 44	16.4	86.3	606	6	CB879649 CB879649
C 45	16.4	86.3	659	4	BJ227679 BJ227679

#### ALIGNMENTS

RESULT 1  
BP055594/c 495 bp mRNA linear EST 25-AUG-2004  
LOCUS BP055594 Lotuscorniculatus var. japonicus pods (less than 20 mm in length) Lotuscorniculatus var. japonicus cDNA clone SPDL057902\_f  
DEFINITION 3', mRNA sequence.

ACCESSION BP055594.1 GI:45612345  
VERSION BP055594  
KEYWORDS EST.  
SOURCE Lotuscorniculatus var. japonicus (Lotuscorniculatus)  
ORGANISM Lotuscorniculatus var. japonicus  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.

REFERENCE 1 (bases 1 to 495)  
AUTHORS Asanizu, E., Nakamura, Y., Sato, S., and Tabata, S.  
TITLE Characteristics of the Lotuscorniculatus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis  
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)  
CONTACT: Erika Asanizu  
COMMENT The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asanizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source  
1..495  
/organism="Lotuscorniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="SPDL057902\_f"  
/tissue\_type="pods"  
/dev\_stage="pod (less than 20 mm in length)"  
/clone\_1lb="Lotuscorniculatus var. japonicus pods (less than 20 mm in length)"

#### ORIGIN

Query Match 91.6%; Score 17.4; DB 5; Length 495;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGGCTTCGCTGGAAGTGG 19  
|||||  
DB 418 TGGGGCTTCGCTGGAAGTGG 400

```

RESULT_2
AV778474/c      508 bp    mRNA     linear   EST 19-AUG-2004
LOCUS          AV778474.Locus japonicus Poda (20-30 mm in length) Locus corniculatus var. japonicus cDNA clone MPD039c04_f 3', mRNA sequence.
ACCESSION      AV778474
VERSION        AV778474
KEYWORDS       AV778474.1 GI:45402131
SOURCE         EST.
ORGANISM       Locus corniculatus var. japonicus (Locus japonicus) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosoids I; Fabales; Fabaceae; Papilionoideae; Loteeae; Locus.
REFERENCE      1. (bases 1 to 508) Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S. Characteristics of the Locus japonicus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis Plant Mol. Biol. 54 (3), 405-414 (2004)
JOURNAL        Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizue@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/planc/. Location/Qualifiers
FEATURES       source
               1..508
                /organism="Locus corniculatus var. japonicus"
                /mol_type="mRNA"
                /isolate="Miyakojima MG-20"
                /db_xref="taxon:34305"
                /clone="MPD039c04_F"
                /tissue_type="Pods (20-30 mm in length)"
                /clone_id="Locus japonicus Pods (20-30 mm in length)"
                /note="vector: pbunescriptii SK-, Site_1: RcorI; Site_2: XhoI"
ORIGIN
Query Match           91.6%; Score 17.4; DB 1; Length 508;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY              1 TGGGTTCTCGTGAAGTCG 19
Db              506 TGGGTTCTCGTGAAGTCG 488
RESULT_3
BP028473            512 bp    mRNA     linear   EST 19-AUG-2004
LOCUS             BP028473.Locus corniculatus var. japonicus flower Locus corniculatus var. japonicus cDNA clone MF005a06_f 3', mRNA sequence.
ACCESSION        BP028473
VERSION          BP028473
KEYWORDS         BP028473.1 GI:45405633
SOURCE           LOTUS.
ORGANISM         Locus corniculatus var. japonicus (Locus japonicus) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosoids I; Fabales; Fabaceae; Papilionoideae; Loteeae; Locus.
REFERENCE        1. (bases 1 to 512) Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S. Characteristics of the Locus japonicus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis Plant Mol. Biol. 54 (3), 405-414 (2004)
JOURNAL          Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizue@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/planc/. Location/Qualifiers
FEATURES

```

```

source
1. .512
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MF05a06_F"
/cisue_type="flower"
/clone_lib="Lotus corniculatus var. japonicus flower"

ORIGIN
Query Match          91.6%; Score 17.4; DB 5; Length 512;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY
1 TGGGGTTGCTGGAAGTGG 19
|||||
505 TGGGGTTGCTGGAAGTGG 487

RESULT 4
AV779844/c
LOCUS
DEFINITION
AV779844 519 bp mRNA linear EST 19-AUG-2004
AV779844 Lotus japonicus Pods (20-30 mm in length) Lotus
corniculatus var. japonicus cDNA clone MPDL066b04_F 3', mRNA
sequence.
ACCESSION
AV779844
VERSION
AV779844.1 GI:45403495
KEYWORDS
EST.
SOURCE
Lotus corniculatus var. japonicus (Lotus japonicus)
Lotus corniculatus var. japonicus
Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
1 (Bases 1 to 519)
Asamizu,S., Nakamura,Y., Sato,S. and Tabata,S.
Characteristics of the Lotus japonicus Gene Repertoire Deduced from
large-Scale Expressed Sequence Tag (EST) Analysis
Plant Mol. Biol. 54 (3), 405-414 (2004)
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 153-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. .519
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL066b04_F"
/cisue_type="Pods (20-30 mm in length)"
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-; Site_1: ScaRI; Site_2:
XhoI"

ORIGIN
Query Match          91.6%; Score 17.4; DB 1; Length 519;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY
1 TGGGGTTGCTGGAAGTGG 19
|||||
508 TGGGGTTGCTGGAAGTGG 490

RESULT 5
AV779524/c
LOCUS
DEFINITION
AV779524 523 bp mRNA linear EST 19-AUG-2004
AV779524 Lotus japonicus Pods (20-30 mm in length) Lotus
corniculatus var. japonicus cDNA clone MPDL059h06_F 3', mRNA
sequence.
ACCESSION
AV779524

```



VERSION AV779524.1 GI:45403176  
KEYWORDS EST.  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM Lotus corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Legae;  
Lotus.  
REFERENCE 1 (bases 1 to 523)  
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from  
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis  
COMMENT Plant Mol. Biol. 54 (3), 405-414 (2004)  
Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers  
1..523  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="MPDL059b06.f"  
/tissue\_type="Pod" (20-30 mm in length)"  
/clone\_lib="Lotus japonicus Pods (20-30 mm in length)"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI"  
ORIGIN  
Query Match 91.6%; Score 17.4; DB 1; Length 523;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 TGGGGTTGCTGGAAGTGG 19  
Db 509 TGGGGTTGCTGGAAGTGG 491  
RESULT 6  
BI419678 526 bp mRNA linear EST 15-AUG-2001  
LOCUS LJNSTR66der Lotus japonicus nodule library 5 and 7 week-old Lotus  
DEFINITION Corniculatus var. japonicus cDNA 5', mRNA sequence.  
ACCESSION BI419678.1 GI:15190701  
VERSION BI419678  
KEYWORDS EST.  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM Lotus corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Legae;  
Lotus.  
REFERENCE 1 (bases 1 to 526)  
AUTHORS Colebatch, G., Freund, S., Trevaekis, B. and Udvardi, M.  
TITLE Lotus japonicus root nodule ESTs: tools for functional genomics  
JOURNAL Unpublished (2000)  
COMMENT Contact: Udvardi MK  
Molecular Plant Nutrition  
Max Planck Institute of Molecular Plant Physiology  
Am Muehlenberg 1, 14476 Golm, Germany  
Fax: 49 331 587 8250  
Email: udvardi@mpimp-golm.mpg.de  
Seq primer: T7  
High quality sequence sccp: 526.  
Location/Qualifiers  
1..526  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/cultiivar="Gifu (B-129)"  
/db\_xref="taxon:34305"  
/dev\_stage="5 and 7 week-old plants"

/clone\_lib="Lotus japonicus nodule library 5 and 7  
week-old"  
/note="Organ: Nodule; Vector: pSPORT1; Site\_1: SalI;  
Site\_2: NotI; The library was prepared using mRNA  
extracted from nodules of 5 and 7 week-old Lotus plants.  
Nodules were induced by, and contained Mesorhizobium  
strain R7A."  
ORIGIN  
Query Match 91.6%; Score 17.4; DB 4; Length 526;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 TGGGGTTGCTGGAAGTGG 19  
Db 75 TGGGGTTGCTGGAAGTGG 93  
RESULT 7  
AV779426/c  
LOCUS AV779426 528 bp mRNA linear EST 19-AUG-2004  
DEFINITION Lotus japonicus Pods (20-30 mm in length) Lotus  
corniculatus var. japonicus cDNA clone MPDL058b01.f 3', mRNA  
sequence.  
ACCESSION AV779426  
VERSION AV779426  
KEYWORDS EST.  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM Lotus corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Legae;  
Lotus.  
REFERENCE 1 (bases 1 to 528)  
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from  
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis  
COMMENT Plant Mol. Biol. 54 (3), 405-414 (2004)  
Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers  
1..528  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="MPDL058b01.f"  
/tissue\_type="Pod" (20-30 mm in length)"  
/clone\_lib="Lotus japonicus Pods (20-30 mm in length)"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI"  
ORIGIN  
Query Match 91.6%; Score 17.4; DB 1; Length 528;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 TGGGGTTGCTGGAAGTGG 19  
Db 509 TGGGGTTGCTGGAAGTGG 491  
RESULT 8  
AV778725/c  
LOCUS AV778725 531 bp mRNA linear EST 19-AUG-2004  
DEFINITION Lotus japonicus Pods (20-30 mm in length) Lotus  
corniculatus var. japonicus cDNA clone MPDL04402.f 3', mRNA  
sequence.  
ACCESSION AV778725  
VERSION AV778725.1 GI:45402382

KEYWORDS EST.  
SOURCE LOTUS corniculatus var. japonicus (Lotus japonicus)  
ORGANISM LOTUS corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.

REFERENCE 1 (bases 1 to 531)  
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis  
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)  
CONTACT: Erika Asamizu  
COMMENT The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers

FEATURES  
source  
1..531  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="MPDL044d02.f"  
/cissue\_type="Pode" (20-30 mm in length)"  
/clone\_1ib="Lotus japonicus Pode (20-30 mm in length)"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

ORIGIN  
Query Match 91.6%; Score 17.4; DB 1; Length 531;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19  
|||||  
506 TGGGGTTCGCTGGAAGTGG 488

Db 506 TGGGGTTCGCTGGAAGTGG 488

RESULT 9 BP043328 532 bp mRNA linear EST 19-AUG-2004  
LOCUS BP043328 Lotus corniculatus var. japonicus flower bud Lotus  
DEFINITION corniculatus var. japonicus cDNA clone MPBL041c10\_f 3', mRNA  
sequence.  
ACCESSION BP043328  
VERSION BP043328.1 GI:45577097  
KEYWORDS EST.  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM LOTUS corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.

REFERENCE 1 (bases 1 to 532)  
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis  
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)  
CONTACT: Erika Asamizu  
COMMENT The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers

FEATURES  
source  
1..532  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="MPBL041c10.f"  
/cissue\_type="flower bud"  
/clone\_1ib="Lotus corniculatus var. japonicus flower bud"

ORIGIN  
Query Match 91.6%; Score 17.4; DB 5; Length 532;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19  
|||||  
509 TGGGGTTCGCTGGAAGTGG 491

Db 509 TGGGGTTCGCTGGAAGTGG 491

RESULT 10 AV777359/c 536 bp mRNA linear EST 19-AUG-2004  
LOCUS AV777359 Lotus japonicus Pode (20-30 mm in length) Lotus  
DEFINITION corniculatus var. japonicus cDNA clone MPDL017e09\_f 3', mRNA  
sequence.  
ACCESSION AV777359  
VERSION AV777359.1 GI:45401024  
KEYWORDS EST.  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM LOTUS corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.

REFERENCE 1 (bases 1 to 536)  
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis  
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)  
CONTACT: Erika Asamizu  
COMMENT The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers

FEATURES  
source  
1..536  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="MPDL017e09.f"  
/cissue\_type="Pode" (20-30 mm in length)"  
/clone\_1ib="Lotus japonicus Pode (20-30 mm in length)"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

ORIGIN  
Query Match 91.6%; Score 17.4; DB 1; Length 536;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19  
|||||  
508 TGGGGTTCGCTGGAAGTGG 490

Db 508 TGGGGTTCGCTGGAAGTGG 490

RESULT 11 AV777186 552 bp mRNA linear EST 19-AUG-2004  
LOCUS AV777186 Lotus japonicus Pode (20-30 mm in length) Lotus  
DEFINITION corniculatus var. japonicus cDNA clone MPDL013e05\_f 3', mRNA  
sequence.  
ACCESSION AV777186  
VERSION AV777186.1 GI:45400851  
KEYWORDS EST.  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM LOTUS corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.

REFERENCE 1 (bases 1 to 552)  
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from  
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis  
CONTACT: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
LOCATION/Qualifiers  
1. 552  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="SPDL053c10\_f"  
/issue\_type="Pods (20-30 mm in length)"  
/clone\_1ib="Lotus japonicus Pods (20-30 mm in length)"  
/note="Vector: pBluescriptII SK-, Site\_1: EcoRI; Site\_2:  
XhoI"

ORIGIN  
Query Match 91.6%; Score 17.4; DB 1; Length 552;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGTTGCTGGAAGTGG 19  
|||||  
512 TGGCGTTGCTGGAAGTGG 494

RESULT 12  
LOCUS BP055162/c 557 bp mRNA linear EST 25-AUG-2004  
DEFINITION BP055162 Lotus corniculatus var. japonicus pods (less than 20 mm in  
length) Lotus corniculatus var. japonicus cDNA clone SPDL050f03\_f  
3', mRNA sequence.  
ACCESSION BP055162  
VERSION BP055162.1 GI:45611913  
KEYWORDS  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM Lotus corniculatus var. japonicus  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eustosida I; Fabales; Fabaceae; Papilionoideae; Legaeae;  
Lotus.  
1 (bases 1 to 557)  
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
Characteristics of the Lotus japonicus Gene Repertoire Deduced from  
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis  
CONTACT: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
LOCATION/Qualifiers  
1. 557  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="SPDL050f03\_f"  
/issue\_type="Pods (less than 20 mm in length)"  
/dev\_stage="pod (less than 20 mm in length)"  
/clone\_1ib="Lotus corniculatus var. japonicus pods (less  
than 20 mm in length)"

ORIGIN  
Query Match 91.6%; Score 17.4; DB 5; Length 557;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGTTGCTGGAAGTGG 19  
|||||  
478 TGGCGTTGCTGGAAGTGG 460

RESULT 13  
LOCUS BP055337/c 565 bp mRNA linear EST 25-AUG-2004  
DEFINITION BP055337 Lotus corniculatus var. japonicus pods (less than 20 mm in  
length) Lotus corniculatus var. japonicus cDNA clone SPDL053c10\_f  
3', mRNA sequence.  
ACCESSION BP055337  
VERSION BP055337.1 GI:45612088  
KEYWORDS  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM Lotus corniculatus var. japonicus  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eustosida I; Fabales; Fabaceae; Papilionoideae; Legaeae;  
Lotus.  
1 (bases 1 to 565)  
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
Characteristics of the Lotus japonicus Gene Repertoire Deduced from  
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis  
CONTACT: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
LOCATION/Qualifiers  
1. 565  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="SPDL053c10\_f"  
/issue\_type="Pods (less than 20 mm in length)"  
/dev\_stage="pod (less than 20 mm in length)"  
/clone\_1ib="Lotus corniculatus var. japonicus pods (less  
than 20 mm in length)"

ORIGIN  
Query Match 91.6%; Score 17.4; DB 5; Length 565;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGTTGCTGGAAGTGG 19  
|||||  
456 TGGCGTTGCTGGAAGTGG 438

RESULT 14  
LOCUS BP056252/c 565 bp mRNA linear EST 25-AUG-2004  
DEFINITION BP056252 Lotus corniculatus var. japonicus pods (less than 20 mm in  
length) Lotus corniculatus var. japonicus cDNA clone SPDL068h12\_f  
3', mRNA sequence.  
ACCESSION BP056252  
VERSION BP056252.1 GI:45613003  
KEYWORDS  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM Lotus corniculatus var. japonicus  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eustosida I; Fabales; Fabaceae; Papilionoideae; Legaeae;  
Lotus.  
1 (bases 1 to 565)  
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
Characteristics of the Lotus japonicus Gene Repertoire Deduced from  
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis  
CONTACT: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
LOCATION/Qualifiers  
1. 565  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="SPDL068h12\_f"  
/issue\_type="Pods (less than 20 mm in length)"  
/dev\_stage="pod (less than 20 mm in length)"  
/clone\_1ib="Lotus corniculatus var. japonicus pods (less  
than 20 mm in length)"

ORIGIN  
Query Match 91.6%; Score 17.4; DB 5; Length 565;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## COMMENT

Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers

## FEATURES

source

1..565  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="SPDL068h12.f"  
/issue\_type="pods"  
/dev\_stage="pod (less than 20 mm in length)"  
/clone\_1fb="Lotus corniculatus var. japonicus pods (less than 20 mm in length)"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 5; Length 565;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## QY

1 TGGGTTTCGCTGGAAGTGG 19  
|||||  
486 TGGGTTTCGCTGGAAGTGG 468

## DB

RESULT 15  
BP034064 583 bp mRNA linear EST 19-AUG-2004  
LOCUS BP034064 Lotu corniculatus var. japonicus flower bud Lotu  
DEFINITION corniculatus var. japonicus cDNA MFB002907\_f\_3', mRNA  
sequence.

ACCESSION BP034064 GI:45411224  
VERSION BP034064.1  
KEYWORDS EST.  
SOURCE Lotu corniculatus var. japonicus (Lotu japonicus)  
ORGANISM Lotu corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;  
Lotus.

## REFERENCE

1 (bases 1 to 583)  
Asamizu E, Nakamura Y, Sato S, and Tabata S.  
Characteristics of the Lotu japonicus Gene Repertoire Deduced from  
Large-Scale Expressed Sequence Tag (EST) Analysis  
Plant Mol. Biol. 54 (3), 405-414 (2004)

## JOURNAL

Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers

## FEATURES

source

1..583  
/organism="Lotu corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="MFB002907.f"  
/issue\_type="flower bud"  
/clone\_1fb="Lotu corniculatus var. japonicus flower bud"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 5; Length 583;  
Best Local Similarity 94.7%; Pred. No. 5.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## QY

1 TGGGTTTCGCTGGAAGTGG 19  
|||||  
449 TGGGTTTCGCTGGAAGTGG 431

Job time : 1906.81 secs

Search completed: November 7, 2005, 09:18:48

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 19:19:37 ; Search time 57.6915 Seconds  
(without alignments)  
482.163 Million cell updates/sec

Title: US-10-777-131a-5

Perfect score: 17

Sequence: 1 agccgcgtagtaacct 17

Scoring table: IDENTITY\_NUC

Searched: Gapop 10.0, Gapext 1.0

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_NA:  
1: /cgn2\_6/prodata/1/ina/5A COMB. seq.\*  
2: /cgn2\_6/prodata/1/ina/5B COMB. seq.\*  
3: /cgn2\_6/prodata/1/ina/5A COMB. seq.\*  
4: /cgn2\_6/prodata/1/ina/5B COMB. seq.\*  
5: /cgn2\_6/prodata/1/ina/5A COMB. seq.\*  
6: /cgn2\_6/prodata/1/ina/5B COMB. seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appl1
C 2	17	100.0	441529	3 US-09-103-840A-1	Sequence 1, Appl1
C 3	14	82.4	246	4 US-09-583-110-1035	Sequence 1035, Ap
C 4	14	82.4	246	4 US-09-583-110-1035	Sequence 1760, Ap
C 5	14	82.4	360	4 US-09-107-433-143	Sequence 143, Appl
C 6	14	82.4	1385	4 US-08-932-411A-17	Sequence 17, Appl
C 7	14	82.4	1385	4 US-09-585-645A-30	Sequence 30, Appl
C 8	14	82.4	1515	4 US-09-726-774-9	Sequence 9, Appl1
C 9	14	82.4	1522	4 US-09-548-606-1	Sequence 1, Appl1
C 10	14	82.4	6109	4 US-09-795-061-1	Sequence 20, Appl
C 11	14	82.4	21338	3 US-08-961-527-20	Sequence 121, Appl
C 12	13.8	81.2	793	2 US-08-967-101-121	Sequence 121, Appl
C 13	13.8	81.2	793	2 US-08-967-101-121	Sequence 121, Appl
C 14	13.8	81.2	793	3 US-08-967-101-121	Sequence 121, Appl
C 15	13.8	81.2	793	3 US-09-124-698-121	Sequence 121, Appl
C 16	13.8	81.2	793	3 US-09-124-698-121	Sequence 121, Appl
C 17	13.8	81.2	793	3 US-08-932-411A-17	Sequence 121, Appl
C 18	13.8	81.2	793	3 US-08-932-411A-17	Sequence 121, Appl
C 19	13.8	81.2	793	4 US-09-583-110-1035	Sequence 121, Appl
C 20	13.8	81.2	793	4 US-08-932-411A-17	Sequence 121, Appl
C 21	13.8	81.2	793	4 US-08-932-411A-17	Sequence 121, Appl
C 22	13.8	81.2	793	4 US-08-932-411A-17	Sequence 121, Appl
C 23	13.8	81.2	793	4 US-08-932-411A-17	Sequence 121, Appl
C 24	13.8	81.2	793	4 US-08-932-411A-17	Sequence 121, Appl
C 25	13.8	81.2	793	4 US-08-932-411A-17	Sequence 121, Appl
C 26	13.8	81.2	793	4 US-08-932-411A-17	Sequence 121, Appl
C 27	13.8	81.2	793	4 US-08-932-411A-17	Sequence 121, Appl

C 28	13.8	81.2	203475	4 US-09-949-016-17228	Sequence 17228, A
C 29	13.8	81.2	203475	4 US-09-949-016-17229	Sequence 17229, A
C 30	13.4	78.8	969	4 US-09-252-991A-2630	Sequence 2630, Ap
C 31	13.4	78.8	1107	4 US-09-252-991A-2281	Sequence 2281, Ap
C 32	13.4	78.8	1596	4 US-09-489-039A-2610	Sequence 2610, Ap
C 33	13.4	78.8	2093	4 US-09-722-971-5	Sequence 7, Appl1
C 34	13.4	78.8	2943	1 US-08-042-747A-7	Sequence 7, Appl1
C 35	13.4	78.8	10059	4 US-09-949-016-15323	Sequence 15323, A
C 36	13	76.5	601	4 US-09-949-016-15323	Sequence 15323, A
C 37	13	76.5	8764	4 US-09-949-016-16041	Sequence 16041, A
C 38	12.8	75.3	51	4 US-09-443-199C-941	Sequence 941, Appl
C 39	12.8	75.3	51	4 US-09-443-199C-941	Sequence 942, Appl
C 40	12.8	75.3	52	3 US-09-291-874-27	Sequence 27, Appl
C 41	12.8	75.3	52	4 US-09-344-783C-44	Sequence 44, Appl
C 42	12.8	75.3	123	4 US-09-902-540-2904	Sequence 2904, Ap
C 43	12.8	75.3	378	4 US-09-252-991A-13252	Sequence 13252, A
C 44	12.8	75.3	471	4 US-09-902-540-6463	Sequence 6463, Ap
C 45	12.8	75.3	582	4 US-09-902-540-4043	Sequence 4043, Ap

## ALIGNMENTS

```
RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
Query Match 100.0%; Score 17; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCCGCGTAGTAACCT 17
Db 1478734 AGCCGCGTAGTAACCT 1478718
RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```

/ SOFTWARE: PatentIn Ver. 2.1  
/ SEQ ID NO 1  
/ LENGTH: 4411529  
/ TYPE: DNA  
/ ORGANISM: Mycobacterium tuberculosis  
/ OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 100.0%; Score 17; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGGTAGTAACCT 17  
DB 1479210 AGCGGTAGTAACCT 1479194

RESULT 3  
US-09-583-110-1035/c  
/ Sequence 1035, Application US/09583110  
/ Patent No. 6699703  
/ GENERAL INFORMATION:  
/ APPLICANT: Lynn Doucette-Stamm et al.  
/ TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
/ TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics  
/ FILE REFERENCE: PAT00-07A  
/ CURRENT APPLICATION NUMBER: US/09/583,110  
/ PRIOR FILING DATE: 2000-05-26  
/ PRIOR APPLICATION NUMBER: US 09/107,433  
/ PRIOR FILING DATE: 1998-06-30  
/ PRIOR APPLICATION NUMBER: US 60/085,131  
/ PRIOR FILING DATE: 1998-05-12  
/ PRIOR APPLICATION NUMBER: US 60/051,553  
/ PRIOR FILING DATE: 1997-07-02  
/ NUMBER OF SEQ ID NOS: 5322  
/ SEQ ID NO 1035  
/ LENGTH: 246  
/ TYPE: DNA  
/ ORGANISM: Streptococcus pneumoniae  
US-09-583-110-1035

Query Match 82.4%; Score 14; DB 4; Length 246;  
Best Local Similarity 100.0%; Pred. No. 2,6e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGTAACCT 17  
DB 131 CGCGTAGTAACCT 118

RESULT 4  
US-09-583-110-1760/c  
/ Sequence 1760, Application US/09583110  
/ Patent No. 6699703  
/ GENERAL INFORMATION:  
/ APPLICANT: Lynn Doucette-Stamm et al.  
/ TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
/ TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics  
/ FILE REFERENCE: PAT00-07A  
/ CURRENT APPLICATION NUMBER: US/09/583,110  
/ PRIOR FILING DATE: 2000-05-26  
/ PRIOR APPLICATION NUMBER: US 09/107,433  
/ PRIOR FILING DATE: 1998-06-30  
/ PRIOR APPLICATION NUMBER: US 60/085,131  
/ PRIOR FILING DATE: 1998-05-12  
/ PRIOR APPLICATION NUMBER: US 60/051,553  
/ PRIOR FILING DATE: 1997-07-02  
/ NUMBER OF SEQ ID NOS: 5322  
/ SEQ ID NO 1760  
/ LENGTH: 246  
/ TYPE: DNA  
/ ORGANISM: Streptococcus pneumoniae  
US-09-583-110-1760

Query Match 82.4%; Score 14; DB 4; Length 246;  
Best Local Similarity 100.0%; Pred. No. 2,6e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGTAACCT 17  
DB 131 CGCGTAGTAACCT 118

RESULT 5  
US-09-107-433-143/c  
/ Sequence 143, Application US/09107433  
/ Patent No. 6800744  
/ GENERAL INFORMATION:  
/ APPLICANT: Lynn A Doucette-Stamm and David Bush  
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
/ SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN  
/ THERAPEUTICS  
/ NUMBER OF SEQUENCES: 5206  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESS: GENOME THERAPEUTICS CORPORATION  
/ STREET: 100 Beaver Street  
/ CITY: Waltham  
/ STATE: Massachusetts  
/ COUNTRY: USA  
/ ZIP: 02354  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: CD-ROM ISO9660  
/ COMPUTER: <Unknown>  
/ OPERATING SYSTEM: <Unknown>  
/ SOFTWARE: <Unknown>  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/09/107,433  
/ FILING DATE: 30-Jun-1998  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 60/ 085131  
/ FILING DATE: May 12, 1998  
/ APPLICATION NUMBER: 60/051553  
/ FILING DATE: July 2, 1997  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Arinello, Pamela Deneke  
/ REGISTRATION NUMBER: 40,489  
/ REFERENCE/DOCKET NUMBER: GTC-011  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (781)893-5007  
/ TELEFAX: (781)893-8277  
/ INFORMATION FOR SEQ ID NO: 143:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 360 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: double  
/ TOPOLOGY: circular  
/ MOLECULE TYPE: DNA (genomic)  
/ HYPOTHETICAL: NO  
/ ANTI-SENSE: NO  
/ ORIGINAL SOURCE:  
/ ORGANISM: Streptococcus pneumoniae  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (8) LOCATION 1...360  
/ SEQUENCE DESCRIPTION: SEQ ID NO: 143:  
US-09-107-433-143

Query Match 82.4%; Score 14; DB 4; Length 360;  
Best Local Similarity 100.0%; Pred. No. 2,6e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGTAACCT 17  
DB 49 CGCGTAGTAACCT 36

```
RESULT 6
US-08-932-411A-17
; Sequence 17, Application US/08932411A
; Patent No. 6566496
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qifu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,411A
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/772,009
; FILING DATE: 19-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,570
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1385 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULAR TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 382..1170
; US-08-932-411A-17

Query Match      82.4%; Score 14; DB 4; Length 1385;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGCGTAGGTAA 14
      |||
Db      104 AGCCGCGTAGGTAA 117

RESULT 7
US-09-585-645A-30
; Sequence 30, Application US/09585645A
; Patent No. 6838444
; GENERAL INFORMATION:
; APPLICANT: Zoghbi, Huda
; APPLICANT: Bellien, Hugo
; APPLICANT: Bermingham, Neesam
; APPLICANT: Hassan, Beesam
; APPLICANT: Ben-Arie, Nissim
; TITLE OF INVENTION: Compositions and Methods for Therapeutic Use of Atonal-associated
; TITLE OF INVENTION: for Deafness, Osteoarthritis, and Abnormal Cell Proliferation
; FILE REFERENCE: P01899US2
```

```
; CURRENT APPLICATION NUMBER: US/09/585,645A
; CURRENT FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: US 60/137,060
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: US 60/176,993
; PRIOR FILING DATE: 2000-01-19
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: MOUSE
US-09-585-645A-30

Query Match      82.4%; Score 14; DB 4; Length 1385;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGCGTAGGTAA 14
      |||
Db      104 AGCCGCGTAGGTAA 117

RESULT 8
US-09-726-774-9
; Sequence 9, Application US/09726774
; Patent No. 6677153
; GENERAL INFORMATION:
; APPLICANT: Iversen, Patrick L.
; TITLE OF INVENTION: Antisense Antibacterial Method and
; TITLE OF INVENTION: Composition
; FILE REFERENCE: 0450-0032.30
; CURRENT APPLICATION NUMBER: US/09/726,774
; CURRENT FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: US 60/168,150
; PRIOR FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 139
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1515
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-726-774-9

Query Match      82.4%; Score 14; DB 4; Length 1515;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 CGCGTAGTAACT 17
      |||
Db      107 CGCGTAGTAACT 120

RESULT 9
US-09-548-606-1
; Sequence 1, Application US/09548606
; Patent No. 6579711
; GENERAL INFORMATION:
; APPLICANT: GAIER, Walter et al.
; TITLE OF INVENTION: NOVEL LACTIC ACID BACTERIA SPECIES
; FILE REFERENCE: 8265-322
; CURRENT APPLICATION NUMBER: US/09/548,606
; CURRENT FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: PCT/EP98/06636
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 97203245.2
; PRIOR FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1522
; TYPE: DNA
; ORGANISM: Bacteria
```

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1522)  
OTHER INFORMATION: n = A,T,C or G  
US-09-548-606-1

Query Match 82.4%; Score 14; DB 4; Length 1522;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGGTAACCT 17  
|||||  
DB 118 CGCGTAGGTAACCT 131

RESULT 10  
US-09-795-061-1/C  
Sequence 1, Application US/09795061  
Patent No. 6759528  
GENERAL INFORMATION:  
APPLICANT: Greenspan, Daniel S  
APPLICANT: Imamura, Yasutada  
TITLE OF INVENTION: Pro-Alpha 3 (V) Collagen Genes  
FILE REFERENCE: 960296,96781  
CURRENT APPLICATION NUMBER: US/09/795,061  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 1  
LENGTH: 6109  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (82)...(5298)  
US-09-795-061-1

Query Match 82.4%; Score 14; DB 4; Length 6109;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGCGTAGGTAA 14  
|||||  
DB 6022 AGCGCGTAGGTAA 6009

RESULT 11  
US-08-961-527-20  
Sequence 20, Application US/08961527  
Patent No. 642035  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2138 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-20

Query Match 82.4%; Score 14; DB 3; Length 2138;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGGTAACCT 17  
|||||  
DB 20823 CGCGTAGGTAACCT 20836

RESULT 12  
US-08-967-101-121  
Sequence 121, Application US/08967101  
Patent No. 5840540  
GENERAL INFORMATION:  
APPLICANT: ST. GEORGE-HYSLOP, PETER H  
APPLICANT: ROMMERS, JOHANNNA M  
APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street Tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,101  
FILING DATE: 10-NOV-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/592,541  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pletcher, Edmund R.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7100  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-967-101-121

Query Match 81.2%; Score 13.8; DB 2; Length 793;  
Best Local Similarity 88.2%; Pred. No. 3.4e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCGCGTAGGTAACCT 17



Db 7 AGCCGCGTATTACT 23

## RESULT 13

US-08-592-541-121  
Sequence 121, Application US/08592541  
Patent No. 5986054

## GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H  
APPLICANT: ROMMENS, JOHANNA M  
APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street Tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,541  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Pitcher, Edmund R.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

US-08-592-541-121

Query Match 81.2%; Score 13.8; DB 2; Length 793;  
Best Local Similarity 88.2%; Pred. No. 3.4e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCCGCGTATTACT 17  
Db 7 AGCCGCGTATTACT 23

## RESULT 14

US-09-124-698-121  
Sequence 121, Application US/09124698  
Patent No. 6117978

## GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H  
APPLICANT: ROMMENS, JOHANNA M  
APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street Tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/124,698  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/592,541  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pitcher, Edmund R.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

US-09-124-698-121

Query Match 81.2%; Score 13.8; DB 3; Length 793;  
Best Local Similarity 88.2%; Pred. No. 3.4e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCCGCGTATTACT 17  
Db 7 AGCCGCGTATTACT 23

RESULT 15  
US-09-127-480-121  
Sequence 121, Application US/09127480  
Patent No. 6194153

GENERAL INFORMATION:  
APPLICANT: ST. GEORGE-HYSLOP, PETER H  
APPLICANT: ROMMENS, JOHANNA M  
APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street Tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/127,480  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,541  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pitcher, Edmund R.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 base pairs  
TYPE: nucleic acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-127-480-121

```

	Query Match	81.2%	Score 13.8;	Length 793;
	Best Local Similarity	86.2%	Ped. No. 3.4e+02;	
	Matches	15; Conservative	0; Mismatches	2; Indels
				Gaps 0;
QY	1 AGCGGCTAGCAACT	17		
Db	7 AGCGGCTATTAACT	23		

Search completed: November 6, 2005, 23:11:06  
Job time : 77.6915 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:11:16 ; Search time 668.064 Seconds  
(without alignments)  
1233.024 Million cell updates/sec

Title: US-10-777-131A-5

Perfect score: 17

Sequence: 1 agccgcgtagtracact 17

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb\_ba: 2: gb\_bt: 3: gb\_in: 4: gb\_om: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_st: 12: gb\_sy: 13: gb\_un: 14: gb\_vi:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	100.0	17	6 AX770107	AX770107 Sequence
2	17	100.0	2488	6 AX770119	AX770119 Sequence
3	17	100.0	37036	1 MSGY151	AD000018 Mycobacte
4	17	100.0	110000	1 AE000516_14	Continuation (15 o
5	17	100.0	299450	1 BX248338	BX842576 Mycobacte
6	17	100.0	348264	1 BX842576	BX842576 Mycobacte
7	15.4	90.6	1291	1 UBA534629	AJ534629 Unculture
8	15.4	90.6	190199	2 AC120897	AC120897 Rattus no
9	15.4	90.6	221666	2 AC103480	AC103480 Rattus no
10	15.4	88.2	180516	10 AC122442	AC122442 Mus muscu
11	14.4	84.7	245	8 ATH531217	AJ531217 Arabidops
12	14.4	84.7	495	6 CO78165	CO78165 Sequence
13	14.4	84.7	1516	1 AY691535	AY691535 Streptoco
14	14.4	84.7	1952	3 AK112388	AK112388 Clona int
15	14.4	84.7	10323	1 AE015457	AE015457 Shewanell
16	14.4	84.7	13099	1 AE013778	AE013778 Yersinia
17	14.4	84.7	103875	9 AL136223	AL136223 Human DNA
18	14.4	84.7	107247	8 CNS08CBP	AL928752 Oryza sat
19	14.4	84.7	110000	1 BX936398_29	Continuation (30 o

C	20	14.4	84.7	110000	2 AC128511_2	Continuation (3 of
	21	14.4	84.7	128619	8 CNS08CBP	AJ935154 Oryza sat
	22	14.4	84.7	139209	2 AC146975	AC146975 Zea mays
	23	14.4	84.7	142781	2 AC141934	AC141934 Rattus no
	24	14.4	84.7	162092	2 AC140825	AC140825 Homo sapi
	25	14.4	84.7	167162	2 BX927367	BX927367 Danio rer
	26	14.4	84.7	169199	9 AC026463	AC026463 Homo sapi
	27	14.4	84.7	175759	10 AF220294	AF220294 Mus muscu
	28	14.4	84.7	183479	2 AC114907	AC114907 Mus muscu
	29	14.4	84.7	184320	2 AC123290	AC123290 Rattus no
	30	14.4	84.7	185988	2 AC025454	AC025454 Homo sapi
	31	14.4	84.7	189639	9 AC010997	AC010997 Homo sapi
	32	14.4	84.7	193050	1 AJ414149	AJ414149 Yersinia
	33	14.4	84.7	198996	2 AC134845	AC134845 Mus muscu
	34	14.4	84.7	207420	2 AC078884	AC078884 Mus muscu
	35	14.4	84.7	235650	2 AC131623	AC131623 Rattus no
	36	14.4	84.7	236702	2 AC125661	AC125661 Rattus no
	37	14.4	84.7	245515	2 AC151275	AC151275 Mus muscu
	38	14.4	84.7	265677	5 CNS09S8K	BX548154 Clona A24
	39	14.4	84.7	265677	5 CNS09S8K	BX576609 Tetradom
	40	14.4	84.7	280857	2 AC115322	AC115322 Rattus no
	41	14.4	84.7	291326	1 AE017135	AE017135 Yersinia
	42	14.4	84.7	296239	2 AC111864	AC111864 Rattus no
	43	14.4	82.4	200	1 AF073802	AF073802 Streptoco
	44	14.4	82.4	200	1 AF073811	AF073811 Streptoco
	45	14.4	82.4	200	1 AF073812	AF073812 Streptoco

#### ALIGNMENTS

RESULT 1	AX770107	Sequence 5 from Patent WO03016562.	17 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770107	AX770107.1	GI:32437685			
DEFINITION	AX770107	AX770107.1	GI:32437685			
ACCESSION	AX770107	AX770107.1	GI:32437685			
VERSION	AX770107	AX770107.1	GI:32437685			
KEYWORDS	AX770107	AX770107.1	GI:32437685			
SOURCE	AX770107	AX770107.1	GI:32437685			
ORGANISM	AX770107	AX770107.1	GI:32437685			
REFERENCE	AX770107	AX770107.1	GI:32437685			
AUTHORS	AX770107	AX770107.1	GI:32437685			
TITLE	AX770107	AX770107.1	GI:32437685			
JOURNAL	AX770107	AX770107.1	GI:32437685			
FEATURES	AX770107	AX770107.1	GI:32437685			
source	AX770107	AX770107.1	GI:32437685			
ORIGIN	AX770107	AX770107.1	GI:32437685			
Query Match	AX770107	AX770107.1	GI:32437685			
Best Local Similarity	AX770107	AX770107.1	GI:32437685			
Matches	AX770107	AX770107.1	GI:32437685			
Db	AX770107	AX770107.1	GI:32437685			
RESULT 2	AX770119	Sequence 17 from Patent WO03016562.	2488 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770119	AX770119.1	GI:32437697			
DEFINITION	AX770119	AX770119.1	GI:32437697			
ACCESSION	AX770119	AX770119.1	GI:32437697			
VERSION	AX770119	AX770119.1	GI:32437697			
KEYWORDS	AX770119	AX770119.1	GI:32437697			
SOURCE	AX770119	AX770119.1	GI:32437697			
ORGANISM	AX770119	AX770119.1	GI:32437697			

REFERENCE Bacteria; Actinobacteria; Actinomycetales;  
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium  
tuberculosis complex.  
1 Gicquel, B.  
AUTHORS Compositions and methods for detecting multidrug resistant strains  
of *M. tuberculosis* having mutations in genes of the *mtrA* family  
JOURNAL Patent: WO 03016562-A 17 27-FEB-2003;  
INSTITUT PASTEUR (FR)  
FEATURES Location/Qualifiers  
source 1..2488  
/organism="Mycobacterium tuberculosis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1773"  
ORIGIN  
Query Match 100.0%; Score 17; DB 6; Length 2488;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGCGCGTAGTAACCT 17  
|||||  
Db 407 AGCGCGTAGTAACCT 423  
|||||  
RESULT 3 MSGY151 37036 bp DNA linear BCT 10-DEC-1996  
LOCUS Mycobacterium tuberculosis sequence from clone y151.  
DEFINITION AD000018  
ACCESSION AD000018  
VERSION AD000018.1 GI:11717734  
KEYWORDS Mycobacterium tuberculosis  
Mycobacterium tuberculosis  
Bacteria; Actinobacteria; Actinomycetales;  
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium  
tuberculosis complex.  
1 (bases 1 to 37036)  
REFERENCE Du, L.  
AUTHORS Direct Submission  
JOURNAL Submitted (11-OCT-1996) L. Du, Genome Therapeutics Corporation, 100  
Beaver Street, Waltham, MA, USA, 02154 du@critic.com  
COMMENT GSDB:S:1004706.  
FEATURES Location/Qualifiers  
source 1..37036  
/organism="Mycobacterium tuberculosis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1773"  
/clone="y151"  
ORIGIN  
Query Match 100.0%; Score 17; DB 1; Length 37036;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGCGCGTAGTAACCT 17  
|||||  
Db 36091 AGCGCGTAGTAACCT 36107  
|||||  
RESULT 4  
AE000516\_14/c  
WPCOMMENT Sequence split into 44 fragments LOCUS AE000516 Accession AE000516  
Fragment Name Begin End  
AE000516\_00 1 110000  
AE000516\_01 100001 210000  
AE000516\_02 200001 310000  
AE000516\_03 300001 410000  
AE000516\_04 400001 510000  
AE000516\_05 500001 610000  
AE000516\_06 600001 710000  
AE000516\_07 700001 810000

AE000516\_08 800001 910000  
AE000516\_09 900001 1010000  
AE000516\_10 100001 1110000  
AE000516\_11 110001 1210000  
AE000516\_12 120001 1310000  
AE000516\_13 130001 1410000  
AE000516\_14 140001 1510000  
AE000516\_15 150001 1610000  
AE000516\_16 160001 1710000  
AE000516\_17 170001 1810000  
AE000516\_18 180001 1910000  
AE000516\_19 190001 2010000  
AE000516\_20 200001 2110000  
AE000516\_21 210001 2210000  
AE000516\_22 220001 2310000  
AE000516\_23 230001 2410000  
AE000516\_24 240001 2510000  
AE000516\_25 250001 2610000  
AE000516\_26 260001 2710000  
AE000516\_27 270001 2810000  
AE000516\_28 280001 2910000  
AE000516\_29 290001 3010000  
AE000516\_30 300001 3110000  
AE000516\_31 310001 3210000  
AE000516\_32 320001 3310000  
AE000516\_33 330001 3410000  
AE000516\_34 340001 3510000  
AE000516\_35 350001 3610000  
AE000516\_36 360001 3710000  
AE000516\_37 370001 3810000  
AE000516\_38 380001 3910000  
AE000516\_39 390001 4010000  
AE000516\_40 400001 4110000  
AE000516\_41 410001 4210000  
AE000516\_42 420001 4310000  
AE000516\_43 430001 440000  
Continuation (15 of 44) of AE000516 from base 1400001 (AE000516 Mycobacterium tuberculo  
Query Match 100.0%; Score 17; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGCGCGTAGTAACCT 17  
|||||  
Db 78753 AGCGCGTAGTAACCT 78737  
|||||  
RESULT 5  
BX248338 299450 bp DNA linear BCT 06-MAY-2004  
LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment  
DEFINITION 5/14.  
ACCESSION BX248338 BX248333  
VERSION BX248338.1 GI:31617962  
KEYWORDS complete genome.  
SOURCE Mycobacterium bovis AF2122/97  
ORGANISM Mycobacterium bovis AF2122/97  
Bacteria; Actinobacteria; Actinomycetales;  
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium  
tuberculosis complex.  
1  
REFERENCE Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,  
AUTHORS Pryor M., Dutkov S., Grondin S., Lacroix C., Monsempé C., Simon S.,  
Harris B., Atkin R., Doggett J., Mayes R., Keating L.,  
Wheeler P.R., Parkhill J., Barrell B.G., Cole S.T., Gordon S.V. and  
Hewinson G.  
TITLE The complete genome sequence of *Mycobacterium bovis*  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 299450)  
AUTHORS Garnier T.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-2003) Garnier T., Unité de Génétique Moléculaire  
Bactérienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex

15, France. e-mail: tgarntier@pasteur.fr Submitted on behalf of the Mycobacterium bovis sequencing teams, TB Research Group, Veterinary Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. PT4 Annotation, Genopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France. Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

## FEATURES

source

1. 299450

Location/Qualifiers

/organism="Mycobacterium bovis AF2122/97"

/mol\_type="genomic DNA"

/strain="AF2122/97"

gene

/db\_xref="taxon:233413"

CDS

/gene="pks3"

/locus\_tag="Mb1213"

CDS

/gene="pks3"

/locus\_tag="Mb1213"

/EC\_number="2.3.1.-"

/note="Mb1213, pks3, len: 2085 aa. Equivalent to Rv1180 and Rv1181, len: 488 aa and 1582 aa, from Mycobacterium tuberculosis strain H37Rv, (100% identity in 488 aa overlap and 99.9% identity in 1582 aa overlap). Probable polyketide beta-ketocacyl synthase (EC 2.3.1.-), similar to the N-terminus of many polyketide synthases e.g. MCAS MYCB0|Q02251 mycocerotic acid synthase from Mycobacterium bovis (2110 aa), FASTA scores: opt: 2115, E(): 0, (66.5% identity in 472 aa overlap). Also similar to, and same length as P96284|Z83858|MTCY24G1.02 M. tuberculosis (496 aa), FASTA scores: opt: 1424, E(): 0, (50.9% identity in 444 aa overlap). Contains possible signal sequence and P800013 Prokaryotic membrane lipoprotein lipid attachment site, also P800606 Beta-ketocacyl synthases active site. BELONGS TO THE BETA-KETOACYL-ACP SYNTHASES FAMILY. Probable polyketide synthase, similar to many e.g. MCAS MYCB0|Q02251 mycocerotic acid synthase from Mycobacterium bovis (2110 aa), FASTA scores: opt: 3518, E(): 0, (59.7% identity in 1614 aa overlap). Note that this similarity extends upstream of the first initiation codon into the upstream MYV005.16; however the stop codon at the end of MYV005.16 is present in at least 4 independent clones (BAC, cosmid and pUC) from the genome. The two CDS's may represent separate modules of the polyketide synthase. REMARK-M.bovis-M.tuberculosis: In Mycobacterium tuberculosis strain H37Rv, Rv1180 and Rv1181 exist as 2 genes. In Mycobacterium bovis, a single base transversion (a-c) leads to a single product (similar to other organisms)."

/codon\_start=1

/transl\_table=11

/product="PROBABLE POLYKETIDE BETA-KETOACYL SYNTHASE Pks3"

/protein\_id="CAD94074.1"

/db\_xref="GI:31617963"

/db\_xref="UniProt/TREMBL:O7U0G2"

/translation="MRTATRTSAVAVGAMCRGDLPGIDSPQRLMEALLRGDDLVGRIIPA DMWDANVYDPDEGVGRSVSRWGAFLDVGDCDFGLTEREATAIDPQRLLELV SWAELEHAGVDPATLAESQTVGFVJLHGDVLDLSDCGAAGPFTGTSSFAGR VATTGLHGPVAVVDACSGSLTAHVGACSLDGDSDLAGGVVLTLEPRVAVGS LQGLMSPTGRCAHFEADAPVSGEGCVLLKRLPLDPAVDGRIYVATIGTAAYGT RYNTAAPSARQAIAYVQOALAAAGVASTGCVTEAHGCTPGYDPEVASTAAYGT BERPCLTSYKTNFGHQSASGRLKLTILAHGVYPQNLHCRIPDQALDIDLELF VPQANTSPQNTGQPRRAVSSYTGKSTGVHALLLEQAPVSEPAASPELTPPEGATL FVSAATSAQQLHVTARLADWDQNGNAGSRVSRMDLYTLSCRAHRPRTVAVTAS PELSALALDVAGDQIPQPAVGHDRGCVVWVSGQSGMPGKGTGLVAEPFATVS AAMEPIYAESGSESVTEAMSAPQTVSGIDRVQPTIFAYVQALAAALKSGVGRGATIG HSLGEAAAVVAGALSHDGLRIGCRSRMLSGMSASVLEPGQOVLSTLARG ISDVLSVVASPSTVYVAGTOSIRDLVAAKQDYLAREVANDVASHPOVDPIIDE LLEVLAEVDPIAREIPIYSATLMDPRERSFTGEVVENURTYVRAAIVQALKQGY RVQGEIAPRPLLTVAEONAAASLMDPIATILAMRGEQLPFGIRGFVADVHNAGAVD FSVQYDPGRVADAPLPSPMTHRTLMLSRSDSHRSHTGAVOAVHPLLGAVHLLLEPERH

gene

CDS

/gene="papa3"

/locus\_tag="Mb1214"

/EC\_number="8.103"

/note="Mb1214, papa3, len: 472 aa. Equivalent to Rv1182, len: 472 aa, from Mycobacterium tuberculosis strain H37Rv, (100% identity in 472 aa overlap). Probable papa3, conserved polyketide synthase (PKS) associated protein, similar to other Mycobacterium hypothetical proteins e.g. Q49618|U00010 B1170\_C1\_180 from Mycobacterium leprae (471 aa), FASTA scores: opt: 2526, E(): 0, (75.6% identity in 471 aa overlap). Similar to other Mycobacterium tuberculosis hypothetical papa proteins; Rv3824c, Rv3820c, Rv1528c."

/codon\_start=1

/evidence=experimental

/transl\_table=11

/product="PROBABLE CONSERVED POLYKETIDE SYNTHASE ASSOCIATED PROTEIN PAPA3"

/protein\_id="CAD94075.1"

/db\_xref="GI:31617964"

/db\_xref="UniProt/TREMBL:O7U0G1"

/translation="MLRVGPIYLTGLDPAASGTSVSRPSVAHTKASQAPISDP VYMQAQRIGYCEQAKGLDYSRLVWVCCQGCQCDIPAANYINAHLRHPTYSM FOYNGGQIIRTIQDPADIEFVPHHGEITLPIQIEIYONTPDPIOMWGCFRGIVG CHFFPEPVDVHVDAMTVGTIMEFHMVAVLVGHAPLIEIPAPGSDDDPCRRQHT FESYTVSEPVQRYANTKRAEGTNGSPDPPLPLGDSKSDADIVYVMHDEQTAOF ESVCTAGARFTGGVLAACGLABHELTGTTTYGLTPRDTRTPADAMTQGFGLIP ETVPVAGARFGDPAARAQTSFDSGVLAADVDRVLESLTLMPPNPVNPVPLDAG AAPLSVLTAEELGTNIQVSDGRYSYQLSIYIVVEQCTAVAVMPPNDPIAEVSAR VIATLKSIVFORVAESGOQNVVA"

/codon\_start=1

/transl\_table=11

/product="PROBABLE CONSERVED POLYKETIDE SYNTHASE ASSOCIATED PROTEIN PAPA3"

/protein\_id="CAD94076.1"

/db\_xref="GI:31617965"

/db\_xref="UniProt/TREMBL:O7U0G1"

/translation="MLRVGPIYLTGLDPAASGTSVSRPSVAHTKASQAPISDP VYMQAQRIGYCEQAKGLDYSRLVWVCCQGCQCDIPAANYINAHLRHPTYSM FOYNGGQIIRTIQDPADIEFVPHHGEITLPIQIEIYONTPDPIOMWGCFRGIVG CHFFPEPVDVHVDAMTVGTIMEFHMVAVLVGHAPLIEIPAPGSDDDPCRRQHT FESYTVSEPVQRYANTKRAEGTNGSPDPPLPLGDSKSDADIVYVMHDEQTAOF ESVCTAGARFTGGVLAACGLABHELTGTTTYGLTPRDTRTPADAMTQGFGLIP ETVPVAGARFGDPAARAQTSFDSGVLAADVDRVLESLTLMPPNPVNPVPLDAG AAPLSVLTAEELGTNIQVSDGRYSYQLSIYIVVEQCTAVAVMPPNDPIAEVSAR VIATLKSIVFORVAESGOQNVVA"

/codon\_start=1

/transl\_table=11

/product="PROBABLE CONSERVED POLYKETIDE SYNTHASE ASSOCIATED PROTEIN PAPA3"

/protein\_id="CAD94077.1"

/db\_xref="GI:31617966"

/db\_xref="UniProt/TREMBL:O7U0G1"

/translation="MLRVGPIYLTGLDPAASGTSVSRPSVAHTKASQAPISDP VYMQAQRIGYCEQAKGLDYSRLVWVCCQGCQCDIPAANYINAHLRHPTYSM FOYNGGQIIRTIQDPADIEFVPHHGEITLPIQIEIYONTPDPIOMWGCFRGIVG CHFFPEPVDVHVDAMTVGTIMEFHMVAVLVGHAPLIEIPAPGSDDDPCRRQHT FESYTVSEPVQRYANTKRAEGTNGSPDPPLPLGDSKSDADIVYVMHDEQTAOF ESVCTAGARFTGGVLAACGLABHELTGTTTYGLTPRDTRTPADAMTQGFGLIP ETVPVAGARFGDPAARAQTSFDSGVLAADVDRVLESLTLMPPNPVNPVPLDAG AAPLSVLTAEELGTNIQVSDGRYSYQLSIYIVVEQCTAVAVMPPNDPIAEVSAR VIATLKSIVFORVAESGOQNVVA"

/codon\_start=1

/transl\_table=11

/product="PROBABLE CONSERVED POLYKETIDE SYNTHASE ASSOCIATED PROTEIN PAPA3"

/protein\_id="CAD94078.1"

/db\_xref="GI:31617967"

/db\_xref="UniProt/TREMBL:O7U0G1"

/translation="MLRVGPIYLTGLDPAASGTSVSRPSVAHTKASQAPISDP VYMQAQRIGYCEQAKGLDYSRLVWVCCQGCQCDIPAANYINAHLRHPTYSM FOYNGGQIIRTIQDPADIEFVPHHGEITLPIQIEIYONTPDPIOMWGCFRGIVG CHFFPEPVDVHVDAMTVGTIMEFHMVAVLVGHAPLIEIPAPGSDDDPCRRQHT FESYTVSEPVQRYANTKRAEGTNGSPDPPLPLGDSKSDADIVYVMHDEQTAOF ESVCTAGARFTGGVLAACGLABHELTGTTTYGLTPRDTRTPADAMTQGFGLIP ETVPVAGARFGDPAARAQTSFDSGVLAADVDRVLESLTLMPPNPVNPVPLDAG AAPLSVLTAEELGTNIQVSDGRYSYQLSIYIVVEQCTAVAVMPPNDPIAEVSAR VIATLKSIVFORVAESGOQNVVA"

/codon\_start=1

/transl\_table=11

/product="PROBABLE CONSERVED POLYKETIDE SYNTHASE ASSOCIATED PROTEIN PAPA3"

/protein\_id="CAD94079.1"

/db\_xref="GI:31617968"

/db\_xref="UniProt/TREMBL:O7U0G1"

/translation="MLRVGPIYLTGLDPAASGTSVSRPSVAHTKASQAPISDP VYMQAQRIGYCEQAKGLDYSRLVWVCCQGCQCDIPAANYINAHLRHPTYSM FOYNGGQIIRTIQDPADIEFVPHHGEITLPIQIEIYONTPDPIOMWGCFRGIVG CHFFPEPVDVHVDAMTVGTIMEFHMVAVLVGHAPLIEIPAPGSDDDPCRRQHT FESYTVSEPVQRYANTKRAEGTNGSPDPPLPLGDSKSDADIVYVMHDEQTAOF ESVCTAGARFTGGVLAACGLABHELTGTTTYGLTPRDTRTPADAMTQGFGLIP ETVPVAGARFGDPAARAQTSFDSGVLAADVDRVLESLTLMPPNPVNPVPLDAG AAPLSVLTAEELGTNIQVSDGRYSYQLSIYIVVEQCTAVAVMPPNDPIAEVSAR VIATLKSIVFORVAESGOQNVVA"

/codon\_start=1

/transl\_table=11

/product="PROBABLE CONSERVED POLYKETIDE SYNTHASE ASSOCIATED PROTEIN PAPA3"

/protein\_id="CAD94080.1"

/db\_xref="GI:31617969"

/db\_xref="UniProt/TREMBL:O7U0G1"

/translation="MLRVGPIYLTGLDPAASGTSVSRPSVAHTKASQAPISDP VYMQAQRIGYCEQAKGLDYSRLVWVCCQGCQCDIPAANYINAHLRHPTYSM FOYNGGQIIRTIQDPADIEFVPHHGEITLPIQIEIYONTPDPIOMWGCFRGIVG CHFFPEPVDVHVDAMTVGTIMEFHMVAVLVGHAPLIEIPAPGSDDDPCRRQHT FESYTVSEPVQRYANTKRAEGTNGSPDPPLPLGDSKSDADIVYVMHDEQTAOF ESVCTAGARFTGGVLAACGLABHELTGTTTYGLTPRDTRTPADAMTQGFGLIP ETVPVAGARFGDPAARAQTSFDSGVLAADVDRVLESLTLMPPNPVNPVPLDAG AAPLSVLTAEELGTNIQVSDGRYSYQLSIYIVVEQCTAVAVMPPNDPIAEVSAR VIATLKSIVFORVAESGOQNVVA"

/codon\_start=1

/transl\_table=11

/product="PROBABLE CONSERVED POLYKETIDE SYNTHASE ASSOCIATED PROTEIN PAPA3"

/protein\_id="CAD94081.1"

/db\_xref="GI:31617970"

/db\_xref="UniProt/TREMBL:O7U0G1"

/translation="MLRVGPIYLTGLDPAASGTSVSRPSVAHTKASQAPISDP VYMQAQRIGYCEQAKGLDYSRLVWVCCQGCQCDIPAANYINAHLRHPTYSM FOYNGGQIIRTIQDPADIEFVPHHGEITLPIQIEIYONTPDPIOMWGCFRGIVG CHFFPEPVDVHVDAMTVGTIMEFHMVAVLVGHAPLIEIPAPGSDDDPCRRQHT FESYTVSEPVQRYANTKRAEGTNGSPDPPLPLGDSKSDADIVYVMHDEQTAOF ESVCTAGARFTGGVLAACGLABHELTGTTTYGLTPRDTRTPADAMTQGFGLIP ETVPVAGARFGDPAARAQTSFDSGVLAADVDRVLESLTLMPPNPVNPVPLDAG AAPLSVLTAEELGTNIQVSDGRYSYQLSIYIVVEQCTAVAVMPPNDPIAEVSAR VIATLKSIVFORVAESGOQNVVA"

PROTEIN MPPL10"  
/protein\_id="CAD94076.1"  
/db\_xref="GI:31617965"  
/db\_xref="GOA:O50439"  
/db\_xref="UniProt/Swiss-Prot:O50439"  
/translation="WGCVNVALVLPVHVPVSLMAQORHVPVVPADAPSVAVROM  
AEAFHSGSENLVVLITDEKGLADENVDRLVDRNDKDVLMQDPTTPELR  
EVLSKDGKAWILPIGLADLGTSPSYHAYTVVERIVKRVAGLTLLVAVGPATVA  
DLTDAGARBASIELAIVMLVILMVYRNVTMLPLVITGASIMTAQALVAVSLV  
VGLASVNOATVLSNMTAGATDYAVFLISYHXYRSGEPERAVORAMSVKVI  
AASATVCTITFGMRPAKIGVSTGPALALIGIYPLAVALLPALIVLSPRGVA  
RGERMATFMRAKGRIRYRPRAYIGASLIGVALASCAIAHFVYDRKOLPEPDS  
SVNGTLEQARATYQAGVGNRLGASRMIDERTDNLRLASGANLADNDVARGQV  
SRAVAGSLVVALAVIQNPGKNTFENIDNARLVSTHIALGALOVNFGIANSF  
DLDSVVALDTSFVDCSNPGCNARVCPHKLQIATDNGTLDKVVGLAQOLSTSPQ  
TVSAVNDGSRISVYRSLKSLGLNDPAAARALISONGANDLASARQVADGVOM  
LVDQTNMGITGANOASAPFLMANGNDASQSMAGFVPPQVLSSEFKKYAOAFISPDG  
HTVRVFIQTDLPFSTAAVDQVNTIITDKAQOPTSLDASISMSGYVMLRDIDY  
YERDMRLIVAVTVVVLITLMLALRIVAPLVGVSVISMSAIGLVAVVQVGLQ  
ELHMSVGLAFVVLVAVGADYMLLASRLDSALGRSVVIRVACTGVTAGGLI  
PLAASMSGLFSSIGTVCGFTIIGVILIDTFVVRITVPAAMATLLGRASWMPGHQ  
RPAPEGQMSAMSARTKTVPOAVDGSKR"  
complement(11182..12261)  
/locus\_tag="Mb1216c"  
complement(11182..12261)  
/locus\_tag="Mb1216c"  
/note="Mb1216c", len: 359 aa. Equivalent to Rv1184c,  
len: 359 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100% identity in 359 aa overlap). Possible exported

Query Match 100.0%; Score 17; DB 1; Length 299450;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AGCCGGCTAGCTAAGCT 17  
DB 162647 AGCCGGCTAGCTAAGCT 162631  
RESULT 6  
BX842576/c 348264 bp DNA linear BCF 10-JUN-2004  
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 5/13.  
DEFINITION BX842576 AL021006 AL021184 AL022000 AL123456 Z73419 Z73902 Z75555  
ACCESSION Z77137 Z77164 Z77826 Z79701 Z80108 Z81011 Z83862 Z95844  
VERSION BX842576.1 GI:38490250  
KEYWORDS complete genome.  
SOURCE Mycobacterium tuberculosis H37Rv  
ORGANISM Mycobacterium tuberculosis H37Rv  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
REFERENCE 1  
AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,  
Tekle, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,  
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
Squares, R., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and  
Barrell, B.G.  
TITLE Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence  
JOURNAL Nature 393 (6685), 537-544 (1998)  
MEDLINE 98295987  
PUBMED 9634230  
REFERENCE 2  
AUTHORS Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.  
TITLE Re-annotation of the genome sequence of Mycobacterium tuberculosis  
H37Rv  
JOURNAL Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)

MEDLINE 22255591  
PUBMED 12368430  
REFERENCES 3 (bases 1 to 348264)  
AUTHORS Parkhill, J.  
TITLE Direct Submission  
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique  
Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux,  
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
On or before Nov 21, 2003 This sequence version replaced  
gi:3242277, gi:3261498, gi:3261541, gi:3261573, gi:3261576,  
gi:3261608, gi:3261593, gi:3261615, gi:3261623, gi:3261635,  
gi:3256012, gi:3242274, gi:3261684, gi:3250713.  
NOTES: Details of M. tuberculosis sequencing at the Sanger Centre are  
available on the world wide web.  
(URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)).

FEATURES  
source  
1..348264  
/organism="Mycobacterium tuberculosis H37Rv"  
/mol\_type="genomic DNA"  
/strain="H37Rv"  
/db\_xref="taxon:83332"  
170..1159  
/gene="mdh"  
/locus\_tag="Rv1240"  
170..1159  
/gene="mdh"  
/locus\_tag="Rv1240"  
/EC\_number="1.1.1.37"  
/function="INVOLVED IN THE CONVERSION OF MALATE TO  
OXALACETATE [CATALYTIC ACTIVITY: (S)-malate + NAD+ =  
oxalacetate + NADH]."  
/note="Rv1240, (MTV006.12), len: 329 aa. Probable mdh,  
malate dehydrogenase (EC 1.1.1.37). Most similar to  
P50917|MDH MYCIE MALATE DEHYDROGENASE from Mycobacterium  
leprae (325 aa). FASTA scores: opt: 1887, E(): 0, (89.1%  
identity in 329 aa overlap). Contains P500068 Malate  
dehydrogenase active site signature. BELONGS TO THE LDH  
FAMILY. MDH SUBFAMILY. Tbpase score is 0.867."  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="PROBABLE MALATE DEHYDROGENASE MDH"  
/protein\_id="CAA15896.1"  
/db\_xref="GI:2695826"  
/translation="MSASPLKVAVTGAQGIYSLFRLASGSLAPDRRIELRLIEI  
EPALQALEGVMLDCAFPPLSGVEIGSPDKIFGVSTALLVGARPRGAGERSDL  
LEANGAIFPAQKALNVAADVVRGVGTNPATNALIAMTNAPDIPIRPFASALTRLD  
HNRAISQIAKGAATVDIKKMTWGNHSATQYDPLFAEVAKNAAEVVDQAWIED  
EPIPTAKKGAALIIDRAGASASASATIDARMDLIGRPADQWMSAVVSGSYCV  
PEGLISSFPVTTGKGMWTVISGLEIDFBRGRIDKTAELADRSATVELGLI"  
635..673  
/gene="mdh"  
/locus\_tag="Rv1240"  
/note="P500068 Malate dehydrogenase active site signature"  
1235..1495  
/locus\_tag="Rv1241"  
1235..1495  
/locus\_tag="Rv1241"  
/locus\_tag="Rv1241"  
/function="UNKNOWN"  
/note="Rv1241, (MTV006.13), len: 86 aa. Conserved  
hypothetical protein, member of family of 16 hypothetical  
Mycobacterium tuberculosis proteins including: PROTEIN C12  
Rv2871|Q10799|YS71 MYCTU HYPOTHETICAL 13.2 kDa  
(124 aa), FASTA scores: opt: 172, E(): 9.5e-06, (37.2%  
identity in 86 aa overlap); Rv2132, Rv3321c, etc. Tbpase  
score is 0.875."  
/codon\_start=1  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAA15897.1"

gene  
/db xref="GI:2659827"  
/translation="MRTTITLDDDVRLVEDVAHRRRPQVINDALRRALAPPVKK  
QOQYRLPHESSAVRSGLDAGFNKLADELEADLADATRRAR"  
1492..1923  
/locus\_tag="RV1242"  
1492..1923  
/locus\_tag="RV1242"  
/locus\_tag="RV1242"  
/function="UNKNOWN"  
/note="RV1242, (MTV006.14), len: 143 aa. Conserved  
hypothetical protein, member of family of 14 hypothetical  
mycobacterial tuberculosis proteins including:  
Rv2872|Q10800|Y572\_MYCTU (147 aa), FASTA scores: opt: 226,  
E(): 2.7e-09, (32.1% identity in 137 aa overlap); Rv0749,  
Rv0277c, Rv02530c, etc. Tbpase score is 0.893."  
/codon\_start=1  
/transl\_table=1  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAAI589.1"  
/db xref="GI:2659828"  
/translation="MIPDINLLYAIVTGPQHRRAHAWMODTVNGHTRIGLTYPALP  
FPLRATATRVIALAPLPADAIAYRMLSPQNDLLTAGRPDIADLGLDKLGNRA  
SHLTVDVLAAGIETVDARHISDTPFARFADIKMTDLRE"  
complement(1946..3634)  
/gene="PE\_PGRS23"  
/locus\_tag="RV1243c"  
complement(1946..3634)  
/gene="PE\_PGRS23"  
/locus\_tag="RV1243c"  
/function="UNKNOWN"  
/note="RV1243c, (MTV006.15c), len: 562 aa. Member of the  
Mycobacterium tuberculosis PE family, PGSS subfamily of  
gly-rich proteins (see Brennan & Delogu 2002). Tbpase  
score is 0.875."  
/codon\_start=1  
/evidence=experimental  
/transl\_table=1  
/product="PE-PEGS FAMILY PROTEIN"  
/protein\_id="CAE55374.1"  
/db xref="GI:38490251"  
/translation="MEVLIADQVLYAAADLEIGISLAALANRAEAPPTGLLAAGA  
DEVSAIPLIFSGNAOAYOALSAOAAAFHOQFVRLSSAGSVAAEAANSPQATV  
DVNSAPITLLIGRLPIIGGANGPGONGDGGILYNGNGGSSSTGCGRGGNGAL  
LIGNGAGAGSPRANGAGANGCWLYNGELGNGCAATITGNGGNGHGGNGAGL  
GNGAGAGALGALGANGQNPVSHOVTATADGCTTGPDENGTDDAGSGSAVAVPGV  
GAGGIGDGDITLQDTVSQAGDEDDGPNFASGAGANGGAASGRGDAVNGNGGA  
GNGGAGGAGGGIGAGGAGSANYANAGNSIGNGAGGANGGIGLAPGAGAGAGNANODP  
PGNSTGNGAGAGDGVGASADVCGAGFGSGSGRGGLILGTGAGAGDGVGDGDI  
GAOGSGGNGNGIGTIGADGMANODGGDGDNGNGDGGAGAGAGVNGNGCTGAGCGL  
GSGSGSGAAGLIGAGANGGAGGGGGCTGPRGAPGDPGTGATGANGQHLNAG"  
3814..4674  
/gene="IPQZ"  
/locus\_tag="RV1244"  
3814..4674  
/gene="IPQZ"  
/locus\_tag="RV1244"  
/gene="IPQZ"  
/locus\_tag="RV1244"  
/function="UNKNOWN"  
/note="RV1244, (MTV006.16), len: 286 aa. Probable  
lipoprotein IPqZ, equivalent tcou15180|MLU1518042 protein  
n1756x from Mycobacterium lepre (228 aa), FASTA scores:  
opt: 1039, E(): 0, (72.5% identity in 229 aa overlap).  
Similar to Mycobacterium tuberculosis hypothetical protein  
Rv3799c. Contains P500013 prokaryotic membrane lipoprotein  
lipid attachment site. Tbpase score is 0.878."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE LIPOPROTEIN IPQZ"  
/protein\_id="CAAI590.1"  
/db xref="GI:2659830"  
/translation="MIRIRILALILAVLIASVAGCSADPGRHPELVYGVSPDSEA  
MLAIAYVALRSYGFARHATAADPAVAKUDSGAPTVVPRACTPMLOTLDPDASVSD  
AOYRAIVSALPEGIAGDITTAABDKPALVVTOSTAKMAGGDLSEPSKCGLLVG  
RVAGHATVAAGPCLPAPREFRNDATMFALRAGQIVAAWTTADADIPADILMTD  
GKPAITREIVTVPLIRRNALTEROLAVNVEVAGVLTDTTALLIGMRQVAAAGDPAVA

```

misc_feature      3850..3882
                  /gene="1pgq2"
                  /locus_tag="RV1244"
                  /note="PS00013 Prokaryotic membrane lipoprotein lipid
                  attachment site"
gene              4755..5585
                  /locus_tag="RV1245C"
                  /complement(4755..5585)
                  /locus_tag="RV1245C"
                  /EC_number="1.-.-.-"
                  /function="UNKNOWN; SUPPOSED INVOLVED IN CELLULAR
                  METABOLISM."
                  /note="RV1245C, (MTV006.17c), len: 276 aa. Probable
                  short-chain dehydrogenase/reductase (EC 1.-.-.-)',
                  equivalent to NP_301801.1|NC_002677 short chain alcohol
                  dehydrogenase from Mycobacterium leprae (277 aa). Also
                  highly similar to various dehydrogenases and
                  oxidoreductases e.g. NP_250228.1|NC_002516 probable
                  short-chain dehydrogenase from Pseudomonas aeruginosa (295
                  aa); NP_421969.1|NC_002696 short chain dehydrogenase
                  family protein from Caulobacter crescentus (278 aa); etc.
                  Also highly similar to others from Mycobacterium

Query Match      100.0%; Score 17; DB 1; Length 348264;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY               1 AGCGCGGTAGGTAACT 17
                  |||||
Db               96169 AGCGCGGTAGGTAACT 96153

RESULT 7
UBA534629        1291 bp      DNA      linear      BCT 19-DEC-2002
LOCUS            Uncultured delta proteobacterium partial 16S rRNA gene, clone
DEFINITION       JG36-T2T-168.
ACCESSION        AJ534629
VERSION          AJ534629.1 GI:26986453
KEYWORDS         16S ribosomal RNA; 16S rRNA gene.
SOURCE           uncultured delta proteobacterium
ORGANISM         Bacteria; Proteobacteria; Deltaproteobacteria; environmental
                 samples.
REFERENCE        1
AUTHORS          Geiseler, A., Tzvetkova, T., Flemming, K. and Selenka-Pobell, S.
TITLE            Comparison of natural bacterial communities found in uranium mining
                 waste piles and mill tailings
JOURNAL          Unpublished
REFERENCE        2 (bases 1 to 1291)
AUTHORS          Selenka-Pobell, S.I.
TITLE            Direct Submission
JOURNAL          Submitted (10-DEC-2002) Selenka-Pobell S.I., Molecular
                 Microbiology, Institute of Radiochemistry, FRG, P.O.B. 51 01 19,
                 D-01314/ Dresden, GERMANY
FEATURES         Location/Qualifiers
                 1..1291
                     /organism="uncultured delta proteobacterium"
                     /mol_type="genomic DNA"
                     /isolation_source="soil sample from uranium mining waste
                     pile near The town of Johannegeorgenstadt"
                     /db_xref="taxon:34034"
                     /clone="JG36-T2T-168"
                     /clone_1lb="16S rDNA (43F-1404R)"
                     /environmental_sample
                     /country="Germany:Saxony"
                     1..1291
                     /gene="16S rRNA"
                     <1..>1291
                     /gene="16S rRNA"
                     /product="16S ribosomal RNA"

```

Query Match 90.6%; Score 15.4; DB 1; Length 1291;  
 Best Local Similarity 94.1%; Pred. No. 1.7e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTACTT 17  
 |||||  
 35 ATCCGCGTAGTACTT 51

RESULT 8  
 AC120897/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AC120897 190199 bp DNA linear HTG 20-NOV-2002  
 Rattus norvegicus clone CH230-267D5, WORKING DRAFT SEQUENCE.  
 AC120897  
 AC120897.4 GI:25137707  
 HTG, HTGS\_PHASE2, HTGS\_DRAFT, HTGS\_FULLTOP.  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 190199)  
 Muzny,D,Marle,M, Metzker,M, Lee, A, Abramson, S, Adams, C, Alder, J, Allen, C, Allen, H, Alsbrooks, S, Amin, A, Anguiano, D, Anyalebechi, V, Aoyagi, A, Ayodeji, M, Baca, B, Baden, H, Baldwin, D, Bandaranaike, D, Barber, M, Barnstead, M, Benahmed, F, Bisswal, K, Blair, J, Blankenburg, K, Blyth, P, Brown, M, Bryant, N, Buhay, C, Burch, P, Burrell, K, Calderon, E, Cardenas, V, Carter, K, Cavazos, I, Caesar, H, Center, A, Chacko, J, Chavez, D, Chen, G, Chen, R, Chen, Y, Chen, Z, Chu, J, Cleveland, C, Cockrell, R, Cox, C, Coyle, M, Cree, A, D'Souza, L, Davila, M, L, Davis, C, Davy-Carroll, L, De Anda, C, Dederich, D, Delgado, O, Denison, S, Deramo, C, Ding, Y, Dinh, H, Diya, K, Draper, H, Dugan-Rocha, S, Dunn, A, Durbin, K, Duval, B, Bayes, K, Egan, A, Escotto, M, Eugene, C, Evans, C, A, Falls, T, Fan, G, Fernandez, S, Finley, M, Flagg, N, Forbes, L, Foster, M, Foster, P, Fraser, C, M, Gabisi, A, Ganta, R, Garcia, A, Garner, T, Garza, M, Gebregreggie, E, Geer, K, Gill, R, Grady, M, Guerra, M, Guevara, W, Gunaratne, P, Haaland, W, Hamill, C, Hamilton, C, Hamilton, K, Harvey, Y, Havlak, P, Hawes, A, Henderson, N, Hernandez, J, Hernandez, R, Hines, S, Hladun, S, L, Hodgson, A, Hogue, M, Hollins, B, Howells, S, Huily, S, Hume, J, Idebrid, D, Jackson, A, Jackson, B, Jacob, L, Jiang, H, Johnson, B, Johnson, R, Jolivet, A, Karpach, S, Kelly, S, Kelly, S, Khan, Z, King, L, Kovar, C, Kowis, C, Kraft, C, L, Lebow, H, Levan, J, Lewis, L, Li, Z, Liu, J, Liu, J, Liu, W, Liu, Y, London, P, Longacre, S, Lopez, J, Lorenshewa, L, Louisedge, H, Lozada, R, J, Lu, X, Ma, J, Maheshwari, M, Mahindaratne, M, Mahmoud, M, Malloy, K, Mangum, A, Mangum, B, Mapua, P, Martin, K, Martin, R, Martinez, B, Mawhinney, S, McLeod, M, P, McNeill, T, Z, Meenen, B, Mloasivile, A, Miner, G, Minja, E, Montemayor, J, Moore, S, Morgan, M, Morris, K, Morris, S, Munidasa, M, Murphy, M, Nair, L, Nankervis, C, Neal, D, Newton, N, Nguyen, A, N, Pal, S, Parks, K, Naoekelen, O, Okwou, G, Olariu, A, A, Pal, S, Parks, K, Pasternak, S, Paul, H, Perez, A, Perez, L, Piamnuch, C, Plopper, F, Polidexter, A, Popovic, D, Primus, B, Pu, L, L, Puzo, M, Quiroz, J, Rachlin, E, Reeves, K, Regier, M, A, Reigh, R, Reilly, B, Reilly, M, Ren, Y, Reuter, M, Richards, S, Riggs, F, Rivers, C, Rodkey, T, Rojce, A, Rose, M, Rose, R, Ruiz, S, J, Sanders, W, Savary, G, Scherer, S, Scott, G, Shateman, S, Shen, H, Shetty, J, Shvartsbeyn, A, Sisson, I, Sitter, C, D, Smaile, D, Sneed, A, Sodergren, E, Song, X, Z, Sorelle, R, Sosa, J, Steidle, M, Strong, R, Sutton, A, Swalek, A, Taboc, P, Taylor, C, Taylor, T, Thomas, N, Thomas, S, Tingey, A, Trejos, Z, Uemari, K, Valas, R, Vera, V, Villaseana, D, Waldron, L, Walker, B, Wang, J, Wang, Q, Wang, S, Warren, J, Warren, R, Wei, X, White, F, Williams, G, Willson, R, Wleczyk, R, Wooden, H, Worley, K, Wright, D, Wright, R, Wu, J, Yakub, S, Yen, J, Yoon, L, Yoon, V, Yu, F, Zhang, J, Zhou, J, Zhou, J, Zhao, S, Zhou, D, von Niederhausern, A, Weiss, R, Smith, D, R, Holt, R, A, Smith, H, O, Weinstock, G, and Gibbs, R, A.

Direct Submission

JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

Unpublished  
 2 (bases 1 to 190199)  
 Worley, K.C.  
 Direct Submission  
 Submitted (14-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 190199)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Nov 20, 2002 this sequence version replaced gi:23614761.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GWT  
 Center clone name: CH230-267D5  
 ----- Summary Statistics  
 Assembly program: PHRAP; version 0.990329  
 Consensus quality: 184736 bases at least Q40  
 Consensus quality: 185958 bases at least Q30  
 Consensus quality: 186579 bases at least Q20  
 Estimated insert size: 190115; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 1 190199: contig of 190199 bp in length.  
 Location/Qualifiers  
 1. 190199  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-267D5"  
 1. 1667  
 /note="wgs end-extension  
 clone\_end:Sp6"  
 2632..3519  
 /note="clone boundary  
 clone\_end:Sp6"  
 site:  
 end\_sequence:BZ168571"  
 187324..188203  
 /note="clone boundary  
 clone\_end:T7"  
 site:  
 end\_sequence:BZ168570"

FEATURES  
 source  
 misc\_feature  
 misc\_feature  
 misc\_feature





clone end:Sp6  
site:EcoRI  
end\_sequence:BH323125"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 2; Length 221666;  
Best Local Similarity 94.1%; Pred. No. 5.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGGCTAGTACT 17  
Db 185623 AGCCGGCTAGTACT 185639

RESULT 10 AC122442 180516 bp DNA 1linear ROD 27-NOV-2003  
AC122442/c Mus musculus BAC clone RP24-230D14 from chromosome 19, complete  
LOCUS DEFINITION sequence.

ACCESSION AC122442  
VERSION AC122442.4 GI:34556312  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 180516)  
Hodges, J. and Bielicki, L.  
The sequence of Mus musculus BAC clone RP24-230D14  
Unpublished (2001)  
2 (bases 1 to 180516)  
Wilson, R.

REFERENCE 2 (bases 1 to 180516)  
Sequencing of Mus musculus  
Unpublished (2001)  
3 (bases 1 to 180516)  
McPherson, J.D. and Waterston, R.H.

REFERENCE 4 (bases 1 to 180516)  
Direct Submission  
Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 180516)  
McPherson, J.D. and Waterston, R.H.

REFERENCE 5 (bases 1 to 180516)  
Direct Submission  
Submitted (01-APR-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
6 (bases 1 to 180516)  
Submitted (10-SEP-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

REFERENCE 6 (bases 1 to 180516)  
Direct Submission  
Submitted (27-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 10, 2003 this sequence version replaced gi:29423952.

REFERENCE 7 (bases 1 to 180516)  
Direct Submission  
Submitted (27-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 10, 2003 this sequence version replaced gi:29423952.

REFERENCE 8 (bases 1 to 180516)  
Direct Submission  
Submitted (27-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 10, 2003 this sequence version replaced gi:29423952.

REFERENCE 9 (bases 1 to 180516)  
Direct Submission  
Submitted (27-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 10, 2003 this sequence version replaced gi:29423952.

REFERENCE 10 (bases 1 to 180516)  
Direct Submission  
Submitted (27-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 10, 2003 this sequence version replaced gi:29423952.

REFERENCE 11 (bases 1 to 180516)  
Direct Submission  
Submitted (27-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 10, 2003 this sequence version replaced gi:29423952.

REFERENCE 12 (bases 1 to 180516)  
Direct Submission  
Submitted (27-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 10, 2003 this sequence version replaced gi:29423952.

REFERENCE 13 (bases 1 to 180516)  
Direct Submission  
Submitted (27-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 10, 2003 this sequence version replaced gi:29423952.

as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

SOURCE INFORMATION:  
The RPI-24 BAC Library has been constructed by Pieter de Jong and  
coworkers (<http://www.choi.org>) from male C57BL/6J mouse spleen  
and/or brain genomic DNA. The clone and detailed information can be  
obtained from Pieter de Jong and coworkers at <http://www.choi.org>

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone.

## FEATURES

## SOURCE

1. 180516  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="19"  
/map="19"  
/clone="RP24-230D14"  
/clone\_1lb="RP24-230D14"  
177..287  
/rpt\_family="A1u"  
repeat\_region  
732..908  
/rpt\_family="B2"  
repeat\_region  
961..1100  
/rpt\_family="B2"  
repeat\_region  
1395..1600  
/rpt\_family="B2"  
repeat\_region  
1703..1891  
/rpt\_family="B2"  
repeat\_region  
2014..2075  
/rpt\_family="A1u"  
repeat\_region  
2175..2378  
/rpt\_family="B2"  
repeat\_region  
2394..2538  
/rpt\_family="A1u"  
repeat\_region  
2946..3048  
/rpt\_family="B4"  
repeat\_region  
3428..3622  
/rpt\_family="B2"  
repeat\_region  
3751..3805  
/rpt\_family="BRY1"  
repeat\_region  
3853..3991  
/rpt\_family="L1"  
repeat\_region  
4004..4076  
/rpt\_family="ID"  
repeat\_region  
4191..4300  
/rpt\_family="A1u"  
repeat\_region  
4403..4596  
/rpt\_family="B2"  
repeat\_region  
4623..4782  
/rpt\_family="B4"  
repeat\_region  
4784..4821  
/rpt\_family="L1"  
repeat\_region  
4892..5083  
/rpt\_family="B2"  
repeat\_region  
5097..5280  
/rpt\_family="B2"  
repeat\_region  
5564..5711  
/rpt\_family="L1"  
repeat\_region  
5837..5919  
/rpt\_family="L1"  
repeat\_region  
5965..6039  
/rpt\_family="A1u"  
repeat\_region  
7028..7162  
/rpt\_family="B2"

```

repeat_region      7344..7456
                    /rpt_family="Alu"
repeat_region      8076..8212
                    /rpt_family="Alu"
repeat_region      8810..8914
                    /rpt_family="Alu"
repeat_region      9356..9875
                    /rpt_family="RMR19B"
repeat_region      9869..10093
                    /rpt_family="RMR19B"
repeat_region      10094..10266
                    /rpt_family="B2"
repeat_region      10267..10403
                    /rpt_family="RMR19B"
repeat_region      10936..11126
                    /rpt_family="B2"
repeat_region      11379..11465
                    /rpt_family="Alu"
repeat_region      14030..14042
                    /rpt_family="D1"
repeat_region      14043..14239
                    /rpt_family="B2"
repeat_region      14240..15016
                    /rpt_family="B1"
repeat_region      15740..15807
                    /rpt_family="ID"
repeat_region      15884..15982
                    /rpt_family="B4"
repeat_region      16454..16513
                    /rpt_family="B4"
repeat_region      16514..16691
                    /rpt_family="B2"
repeat_region      16724..16932
                    /rpt_family="B4"
repeat_region      17027..17241
                    /rpt_family="B2"
repeat_region      17507..17582
                    /rpt_family="ID"
repeat_region      17880..17975
                    /rpt_family="Alu"
repeat_region      17986..18056
                    /rpt_family="ID"
repeat_region      18118..18181
                    /rpt_family="Alu"
repeat_region      18135..18147
                    /note="Sequence derived from one plasmid subclone."
unsure
repeat_region      18295..18330
                    /rpt_family="Alu"
repeat_region      18306..18393
                    /rpt_family="B4"
repeat_region      19340..19363
                    /rpt_family="B2"
repeat_region      19364..19512
                    /rpt_family="Alu"
repeat_region      19513..19680
                    /rpt_family="B2"
repeat_region      19905..20125
                    /rpt_family="B4"
repeat_region      20602..20745
                    /rpt_family="B2"
repeat_region      20915..21176
                    /rpt_family="B4"
repeat_region      21294..21441
                    /rpt_family="Alu"
repeat_region      21520..21714
                    /rpt_family="B4"
repeat_region      21932..22139
                    /rpt_family="B2"
repeat_region      22457..22672
                    /rpt_family="B4"

```

Query Match 88.2%; Score 15; DB 10; Length 180516;  
Best Local Similarity 100.0%; Pred. No. 9.5e+02;

```

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACCCGCTAGTAC 15
    |||||
Db 25935 ACCCGCTAGTAC 25921

RESULT 11
ATHS31217/c 245 bp DNA linear PLN 29-MAR-2003
LOCUS Arabidopsis thaliana T-DNA flanking sequence, left border, clone
228D07.
ACCESSION AJ531217
VERSION AJ531217.1 GI:26799477
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eucosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F.,
          Chauvin, S., Bechtold, N., Crnaud, C., DeRose, R., Pelletier, G.,
          Lepoint, L., Caboche, M. and Lecharny, A.
          T-DNA integration into the Arabidopsis genome depends on sequences
          of pre-insertion sites
          EMBO Rep. 3 (12), 1152-1157 (2002)
JOURNAL MEDLINE
PUBMED 22363535
REFERENCE 2 (bases 1 to 245)
AUTHORS Balzerque, S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue
          Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana
          plants from INRA (Versailles). The DNA fragment(s) resulting from
          the PCR were directly sequenced from the left or the right border
          to determine the genomic sequence flanking the insertion. T-DNA
          derived sequences were removed. Information to order the
          corresponding mutant line and a link to a database providing a
          graphical display of the insertion site are available at
          http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
          been generated in the framework of the French plant genomics
          program 'Genoplante' (http://www.genoplante.com and
          http://genoplante-info.inbio.gen.fr).
FEATURES
source
    1..245
       /organism="Arabidopsis thaliana"
       /mol_type="genomic DNA"
       /culivar="Massillaewek1ja"
       /db_xref="taxon:3702"
       /clone="228D07"
       /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
misc_feature
    1..245
       /note="T-DNA flanking sequence
       left border"
ORIGIN
Query Match 84.7%; Score 14.4; DB 8; Length 245;
Best Local Similarity 93.8%; Pred. No. 8.8e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 GCCGCTAGTACT 17
    |||||
Db 70 GCCGCTAGTACT 55

RESULT 12
CQ738165/c 495 bp DNA linear PAT 03-FEB-2004
LOCUS Sequence 24099 from Patent WO02068579.
DEFINITION CQ738165
ACCESSION CQ738165.1 GI:42338344
VERSION

```

KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE Kite, such as nucleic acid arrays, comprising a majority of humenexons or transcripts, for detecting expression and other uses thereof  
JOURNAL Patent: WO 02068579-A 24099 06-SEP-2002;  
PE Corporation (NY) (US)  
FEATURES  
source Location/Qualifiers  
1. .495  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 84.7%; Score 14.4; DB 6; Length 495;  
Best Local Similarity 93.8%; Pred. No. 7.5e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTACC 16  
|||||  
Db 226 AGCCGCGTAGTACC 211

RESULT 13  
AY691535 1516 bp DNA linear BCT 28-AUG-2004  
LOCUS Streptococcus anginosus strain CHDC Y33 16S ribosomal RNA gene,  
DEFINITION partial sequence.  
ACCESSION AY691535  
VERSION AY691535.1 GI:51493556  
KEYWORDS Streptococcus anginosus  
SOURCE Streptococcus anginosus  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (bases 1 to 1516)  
AUTHORS Kook, J.-K., Yoo, S.-Y., Kim, P.-S., Hwang, H.-K., Choe, S.-J., Kim, D.-K.  
TITLE Identification of Non-mutans Streptococci in Saliva Recovering on  
JOURNAL Multis-Salivarius Bacitracin Agar Medium  
REFERENCE 2 (bases 1 to 1516)  
AUTHORS Kook, J.-K., Yoo, S.-Y., Kim, P.-S., Hwang, H.-K., Choe, S.-J., Kim, D.-K.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUL-2004) Dept. of Oral Biochemistry, Conservative  
dentistry, and Public Health and Preventive Dentistry, College of  
Dentistry, Chosun University, 375 Seo-suk Dong, Dong-gu, Gwang-ju  
501-759, Republic of Korea  
FEATURES  
source Location/Qualifiers  
1. .1516  
/organism="Streptococcus anginosus"  
/mol\_type="genomic DNA"  
/strain="CHDC Y33"  
/db\_xref="taxon:1328"  
<1. .>1516  
/product="16S ribosomal RNA"

ORIGIN  
Query Match 84.7%; Score 14.4; DB 1; Length 1516;  
Best Local Similarity 93.8%; Pred. No. 5.8e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTACCCT 17  
|||||  
Db 118 GCCGCGTAGTACCCT 133

RESULT 14  
AK112388 1952 bp mRNA linear INV 30-NOV-2002  
LOCUS Ciona intestinalis cDNA, clone: ciad046102, full insert sequence.  
DEFINITION AK112388  
ACCESSION AK112388.1 GI:23575439  
VERSION  
KEYWORDS cDNA  
SOURCE Ciona intestinalis  
ORGANISM Ciona intestinalis  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
Phlebobranchia; Clonidae; Ciona.

REFERENCE 1  
AUTHORS Satou, Y., Yamada, L., Mochizuki, Y., Takatori, N., Kawashima, T.,  
Sasaki, A., Hamaguchi, M., Awazu, S., Yagi, K., Sasekura, Y.,  
Nakayama, A., Ishikawa, H., Inaba, K. and Satoh, N.  
TITLE A cDNA resource from the basal chordate Ciona intestinalis  
JOURNAL Genes 33 (4), 153-154 (2002)  
MEDLINE 22191024  
PUBMED 12203911  
REFERENCE 2 (bases 1 to 1952)  
AUTHORS Satou, Y. and Satoh, N.  
TITLE Direct Submission  
JOURNAL Submitted (04-OCT-2002) Nori Satoh, Kyoto University, Department of  
Zoology, Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
(E-mail: satoh@esaidian.zool.kyoto-u.ac.jp, Tel: 81-75-753-4095,  
Fax: 81-75-705-1113)  
COMMENT Ciona intestinalis cDNA Project (URL:  
http://ghost.zool.kyoto-u.ac.jp/index1.html).  
Location/Qualifiers  
1. .1952  
/organism="Ciona intestinalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:7719"  
/clone="ciad046102"

ORIGIN  
Query Match 84.7%; Score 14.4; DB 3; Length 1952;  
Best Local Similarity 93.8%; Pred. No. 5.5e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTACCCT 17  
|||||  
Db 1595 GCCGCGTAGTACCCT 1610

RESULT 15  
AE015457 10323 bp DNA linear BCT 02-DEC-2002  
LOCUS Shewanella oneidensis MR-1 section 6 of 457 of the complete genome.  
DEFINITION AE015457 AE014299  
ACCESSION AE015457.1 GI:24345417  
VERSION  
KEYWORDS  
SOURCE Shewanella oneidensis MR-1  
ORGANISM Shewanella oneidensis MR-1  
Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
Shewanellaceae; Shewanella.

REFERENCE 1 (bases 1 to 10323)  
AUTHORS Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, M.,  
Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R.,  
Meyer, T., Tsaplin, A., Scott, J., Beaman, M., Brinkac, L., Daugherty, S.,  
DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R.,  
Peterson, J., Unayam, L., White, O., Wolf, A., Vamathevan, J.,  
Weldman, J., Impraim, M., Lee, K., Berry, K., Lee, C., Mueller, J.,  
Khoury, H., Gill, J., Utechback, T., McDonald, L., Feldblyum, T.,  
Smith, H., Venter, J., Neilson, K. and Fraser, C.  
TITLE Genome sequence of the dissimilatory metal ion-reducing bacterium  
JOURNAL Shewanella oneidensis  
Nat. Biotechnol. 20 (11), 1118-1123 (2002)  
MEDLINE 22297686  
PUBMED 12368613  
REFERENCE 2 (bases 1 to 10323)  
AUTHORS Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, M.,  
Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R.,

TITLE  
JOURNAL

Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S., DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R., Peterson, J., Umayam, L., White, O., Wolf, A., Yamanevan, J., Weidman, J., Impraim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Kouri, H., Gill, J., Uterback, T., McDonald, L., Feldblum, T., Smith, H., Venter, J., Nealson, K. and Fraser, C.  
Direct Submission  
Submitted (12-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

## FEATURES

## source

Location/Qualifiers  
1. .10323

/organism="Shewanella oneidensis MR-1"

/mol\_type="Genomic DNA"

/strain="MR-1"

/db\_xref="taxon:211586"

124. .597

/gene="secB"

/locus\_tag="SO0052"

124. .597

/gene="secB"

/locus\_tag="SO0052"

/note="Identified by match to TIGR protein family HMM  
TIGR00809"

/codon\_start=1

/product="protein-export protein SecB"

/protein\_id="FAN53139.1"

/db\_xref="GI:24345418"

/translation="MAEVANNEQAPQFNIRVYTKDVSEFPNSPAVFOKEMNPEVK  
LDLDRSKLADVYEVLSLTVQAQSGSTAFCEVOQAGISLIGLTLPOLAHSG  
AYCPNLPFPAAREAVSGSLVGRGTFPQNLAPVPFDALFAQYVQORQAALAEANA"

602. .1618

/gene="gpaA"

/locus\_tag="SO0053"

602. .1618

/gene="gpaA"

/locus\_tag="SO0053"

/note="similar to GB:L06148, SP:008378, PID:553309,  
GB:L06148, SP:Q08378, and PID:553309; identified by  
sequence similarity; putative"

/codon\_start=1

/translation="table=11

/product="glycerol-3-phosphate dehydrogenase (NAD(P)+)"

/protein\_id="FAN53140.1"

/db\_xref="GI:24345419"

/translation="MNNSADITVVGAGSYTALAISLASGHKTLMGHPAHMOTLA  
EDCNQNFLEGLAEPFECLEADLAKALASNNVLVVPSPHVFQVLAQKPLLRDA  
RIWATKGLPEPTEGLQDVARDVLGEQYPLAVLSGTFPAKELAMGLPTISVAGTCD  
QFPAELVELLHSPKRLVYANDDFIGLQGLAVSNVLAIGASDGIQFANARTALI  
TRGLVELTRGLGALGASTAFMGAGLGDVLCTQNSRRRPGALGKGCVDVTAO  
VEIGOVVEGRNKEVFTLAKRLGVEMPTIEQIYQVLYQSKSPVDAAKELIGREKISE  
TPQ"

1968. .3158

/locus\_tag="SO0054"

1968. .3158

/locus\_tag="SO0054"

/note="Identified by match to Pfam protein family HMM  
pf03486"

/codon\_start=1

/translation="table=11

/product="conserved hypothetical protein TIGR00275"

/protein\_id="FAN53141.1"

/db\_xref="GI:24345420"

/translation="MSYKHHVITIGGAAGLCAATAGYGRGDVLYLVNKAQGRKI  
LISGGRCNFTLVKPEPANTCGNPHVNSALARPYSQQTLEVERHGLEYHRRDQ  
LFCNDASKEIVTMLTECEWAGVSIKLRDILSVGNKAGRFELINTSGLSGDSVIT  
ATGGLSMPKGLATPYGOLAEQGLKYLPTHAGLVPTMSEDKIPPEPLSGIAPSR  
ITAKDGTAFSEALPTFHRLGSGPALQIISNYKAGETIKINLPPMDAQAISQOLAA  
HPKQSLRNTSLMLPKRIYEVLPDEVILANKALNOLVHAREAKLYVDLHPVTVMNGRE  
GYTAEVTLGGVDTHELSSKTMIAIKVPLGFIQGVNDVSGMLGGRFPQAMASGVAA  
GRKV"

complement (3381. .5060)

/locus\_tag="SO0055"

/gene="kdpB"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

/locus\_tag="SO0055"

## CDS

complement (3381. .5060)

/locus\_tag="SO0055"

/note="similar to SP:P26167, and PID:995679; identified  
by sequence similarity; putative"

/codon\_start=1

/translation="table=11

/product="conserved domain protein"

/protein\_id="FAN53142.1"

/db\_xref="GI:24345421"

/translation="MRKARDNRPVLFTGAGSKSAGIPLASELVNEINEKALDLK  
ALSATRYDYGRCAALTKDDRRELLQDYIOKARINNAHIVQALLLEKGYORVLTN  
FNIILSRSCGLGLPSIYDFTSANLIDLSLIDPSIVHLHGSGHGFQVQNTETN  
HAELSGFVANTLNQSPSIFIGYSGNADAFPEPLKDKFTPOKRLFWGRSKABHN  
ENLAPNVNHYLHNEHDADEFELIAQELCFPRKADDDYGHLLDLEPVLSEPTQO  
HNELDIASTRSHKRSQEREDARPANELLMKGTNELIALKGTEDLEEKAAV  
AMALIALDLLEQAKKAQPKLFEOSFEKYQALALIKPMHEKALNMGALSDLAOL  
KQEPALFEQSLBKQALALIKPMHDLFWMGVLSLAOLKQEPALFEOSIEKYEA  
LAIKPDHEAISMWGCALDLAKLQOQTELFESKSLIOARSVANOPLYNLAQVSL  
LHDDLNCKEELPACYKAKTLPDKNHNLNTDTDLDPVRNKEPFKELDLTLPQ"

5448. .6206

/locus\_tag="SO0056"

5448. .6206

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

/locus\_tag="SO0056"

## gene

complement (3381. .5060)

## CDS

/locus\_tag="SO0055"

## gene

complement (3381. .5060)

## CDS

/locus\_tag="SO0055"

## gene

complement (3381. .5060)

CDS

/locus\_tag="S00059"  
complement(8425..9114)  
/gene="Kdpe"  
/locus\_tag="S00059"  
/note="similar to SP:002940, and P1D:151433; identified  
by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulatory protein Kdpe"  
/protein\_id="AAN53146.1"  
/db\_xref="GI:24345425"  
/translation="MAVKYLVDDEPOHTFMRISLBAGFEYLSATSIGALKQYRS  
HQPRLIVDLGPPDGDGIELLMRQDQKTPVLVLTARPEEKIRLEAGANDYLSK  
PFGIRSLVIRKRLDVTISIAADVLOFPLKQKSHQCTIANQOIRIATKKEPAL  
LAHLSHPQQLVKQSELLRRIRWGETHQEDGHYLRILVSQLRKLNDSDEQQLKTEPG  
LGYRLTDLGLP"  
complement(9115..10161)  
/locus\_tag="S00060"  
complement(9115..10161)  
/locus\_tag="S00060"  
/note="similar to GP:9947604; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="sensor histidine kinase"  
/protein\_id="AAN53147.1"  
/db\_xref="GI:24345426"  
/translation="MTFANTASINHYMRTHPALFTAMLVVLITSYFIDVFGSAIA

gene

CDS

Query Match 84.7%; Score 14.4; DB 1; Length 10323;  
Best Local Similarity 93.8%; Pred. No. 3.8e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACC 16  
|||||  
Db 2848 AGCCGCTTAGTAACC 2833

Search completed: November 7, 2005, 06:42:15  
Job time : 676.064 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21 / Search time 196.043 Seconds  
(without alignments)  
513.336 Million cell updates/sec

Title: US-10-777-131a-5

Perfect score: 17

Sequence: 1 agcgcgcagtaacct 17

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

N\_Geneseq\_16Dec04:\*  
1: geneseq1980s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001as:\*  
5: geneseq2001bs:\*  
6: geneseq2002as:\*  
7: geneseq2002bs:\*  
8: geneseq2003as:\*  
9: geneseq2003bs:\*  
10: geneseq2003cs:\*  
11: geneseq2003ds:\*  
12: geneseq2004as:\*  
13: geneseq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	100.0	17	8	ABZ79931 Mycobacte
2	17	100.0	2488	8	ABZ79947 Mycobacte
3	17	100.0	110000	4	AA199682_14
4	17	100.0	110000	4	AA199683_14
5	14.4	84.7	505	2	AA21192
6	14.4	84.7	735	8	ACA54363
7	14.4	84.7	2967	5	AA593164
8	14.4	82.4	237	9	ACL21756
9	14.4	82.4	317	9	ACL21751
10	14.4	82.4	360	13	ADR91508
11	14.4	82.4	394	9	ACL21733
12	14.4	82.4	353	9	ACL21732
13	14.4	82.4	561	9	ACL21732
14	14.4	82.4	582	9	ACL21772
15	14.4	82.4	591	9	ACL21729
16	14.4	82.4	596	9	ACL21768
17	14.4	82.4	602	9	ACL21752
18	14.4	82.4	603	9	ACL21770
19	14.4	82.4	605	9	ACL21750
20	14.4	82.4	634	9	ACL21736

21	14	82.4	639	9	ACL21731	ACL21731 DNA clone
22	14	82.4	653	9	ACL21730	ACL21730 DNA clone
23	14	82.4	664	9	ACL21744	ACL21744 DNA clone
24	14	82.4	694	9	ACL21771	ACL21771 DNA clone
25	14	82.4	698	9	ACL21757	ACL21757 DNA clone
26	14	82.4	700	9	ACL21755	ACL21755 DNA clone
27	14	82.4	700	9	ACL21764	ACL21764 DNA clone
28	14	82.4	702	9	ACL21747	ACL21747 DNA clone
29	14	82.4	703	9	ACL21748	ACL21748 DNA clone
30	14	82.4	738	9	ACL21746	ACL21746 DNA clone
31	14	82.4	751	9	ACL21769	ACL21769 DNA clone
32	14	82.4	1299	4	AAF90009	AAF90009 Nucleotide
33	14	82.4	1334	6	AB571611	AB571611 Streptococ
34	14	82.4	1335	10	ADB61679	ADB61679 16S rRNA
35	14	82.4	1385	2	AAV27049	AAV27049 Mouse neu
36	14	82.4	1385	3	AAZ51980	AAZ51980 Murine neu
37	14	82.4	1385	5	AAF27269	AAF27269 Mouse neu
38	14	82.4	1385	5	AAI62216	AAI62216 Mouse ato
39	14	82.4	1470	4	AAF31086	AAF31086 Bacterial
40	14	82.4	1515	5	AA511029	AA511029 Streptococ
41	14	82.4	1515	10	ADB61678	ADB61678 16S rRNA
42	14	82.4	1522	2	AA577721	AA577721 16S rRNA
43	14	82.4	1531	10	ADB61677	ADB61677 16S rRNA
44	14	82.4	6109	4	AAH78667	AAH78667 Murine Co
45	14	82.4	7250	6	ABN72583	ABN72583 Streptococ

## ALIGNMENTS

RESULT 1  
ID ABZ79931 standard; DNA; 17 BP.  
XX AC ABZ79931;  
XX AC 19-MAY-2003 (first entry)  
XX DE Mycobacterium tuberculosis alka-1 PCR primer SEQ ID NO:5.  
XX KM Mycobacterium tuberculosis; mutT2; alka; ogt; RV3908; mutY; RV3909;  
XX KM detection; multidrug resistance; multiple drug resistance; MDR;  
XX KM infection; PCR primer; ss.  
XX OS Mycobacterium tuberculosis.  
XX OS Synthetic.  
XX PN WO2003016562-A2.  
XX PD 27-FEB-2003.  
XX PF 14-AUG-2002; 2002WO-EP009679.  
XX PR 14-AUG-2001; 2001US-0311824P.  
XX PR 21-AUG-2001; 2001US-0313523P.  
XX PA (INSP) INST PASTEUR.  
XX PI Gicquel B;  
XX WPI; 2003-256711/25.  
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate  
PT by detecting an alteration in the DNA repair system of the isolate.  
XX PS Claim 32; Page 16; 83pp; English.  
CC The present invention describes a method for predicting the epidemic  
CC character of a Mycobacterium tuberculosis isolate and/or a selective  
CC advantage to be maintained in the host and/or the acquisition of multiple  
CC drug resistance (MDR) by the isolate comprising detecting an alteration  
CC in the DNA repair system of the isolate. Also described: (1) detecting a

CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polymucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a PCR primer for M. tuberculosis alkA,  
CC which is used in the exemplification of the present invention  
SQ Sequence 17 BP; 4 A; 5 C; 5 G; 3 T; 0 U; 0 Other;

Query Match 100.0%; Score 17; DB 8; Length 17;  
Best Local Similarity 100.0%; Pred. No. 8;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGCGTAGTAACCT 17  
DB 1 AGCGCGTAGTAACCT 17

RESULT 2  
AB279947 standard; DNA; 2488 BP.  
XX AB279947;  
AC AB279947;  
XX 19-MAY-2003 (first entry)  
DT Mycobacterium tuberculosis alkA nucleotide sequence SEQ ID NO:17.  
DE Mycobacterium tuberculosis; mutT2; alkA; ogc; R33908; mutY; R33909;  
XX Mycobacterium tuberculosis; mutT2; alkA; ogc; R33909;  
XX detection; multidrug resistance; multiple drug resistance; MDR;  
XX infection; gene; de.  
OS Mycobacterium tuberculosis.  
XX WO2003016562-A2.  
PN 27-FEB-2003.  
PD 14-AUG-2002; 2002WO-EP009679.  
XX 14-AUG-2001; 2001US-0311824P.  
PR 21-AUG-2001; 2001US-0313523P.  
XX (INSP ) INST PASTEUR.  
PA Gicquel B;  
PI WPI; 2003-256711/25.  
XX Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate  
PT by detecting an alteration in the DNA repair system of the isolate.  
XX Disclosure; Fig 2A-B; 83pp; English.  
PS The present invention describes a method for predicting the epidemic  
CC character of a Mycobacterium tuberculosis isolate and/or a selective  
CC advantage to be maintained in the host and/or the acquisition of multiple  
CC in the DNA repair system of the isolate. Also described: (1) detecting a  
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a  
CC polymucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)  
CC an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)  
CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
CC risk of being unable to eliminate the bacillus or of developing MDR  
CC tuberculosis. The method is useful for predicting the epidemic character  
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
CC be maintained in the host and/or the acquisition of MDR by the isolate.  
CC The present sequence represents a M. tuberculosis alkA nucleotide

CC sequence, which is used in the exemplification of the present invention  
XX SQ Sequence 2488 BP; 426 A; 847 C; 807 G; 408 T; 0 U; 0 Other;

Query Match 100.0%; Score 17; DB 8; Length 2488;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGCGTAGTAACCT 17  
DB 407 AGCGCGTAGTAACCT 423

RESULT 3  
AA199682\_14/C  
Continuation (15 of 45) of AA199682 from base 1400001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments  
WP Sequence split into 45 fragments

Fragment Name	Begin	End
WP AA199682_00	1	110000
WP AA199682_01	100001	210000
WP AA199682_02	200001	310000
WP AA199682_03	300001	410000
WP AA199682_04	400001	510000
WP AA199682_05	500001	610000
WP AA199682_06	600001	710000
WP AA199682_07	700001	810000
WP AA199682_08	800001	910000
WP AA199682_09	900001	1010000
WP AA199682_10	1000001	1110000
WP AA199682_11	1100001	1210000
WP AA199682_12	1200001	1310000
WP AA199682_13	1300001	1410000
WP AA199682_14	1400001	1510000
WP AA199682_15	1500001	1610000
WP AA199682_16	1600001	1710000
WP AA199682_17	1700001	1810000
WP AA199682_18	1800001	1910000
WP AA199682_19	1900001	2010000
WP AA199682_20	2000001	2110000
WP AA199682_21	2100001	2210000
WP AA199682_22	2200001	2310000
WP AA199682_23	2300001	2410000
WP AA199682_24	2400001	2510000
WP AA199682_25	2500001	2610000
WP AA199682_26	2600001	2710000
WP AA199682_27	2700001	2810000
WP AA199682_28	2800001	2910000
WP AA199682_29	2900001	3010000
WP AA199682_30	3000001	3110000
WP AA199682_31	3100001	3210000
WP AA199682_32	3200001	3310000
WP AA199682_33	3300001	3410000
WP AA199682_34	3400001	3510000
WP AA199682_35	3500001	3610000
WP AA199682_36	3600001	3710000
WP AA199682_37	3700001	3810000
WP AA199682_38	3800001	3910000
WP AA199682_39	3900001	4010000
WP AA199682_40	4000001	4110000
WP AA199682_41	4100001	4210000
WP AA199682_42	4200001	4310000
WP AA199682_43	4300001	4410000
WP AA199682_44	4400001	4411529

Query Match 100.0%; Score 17; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGCGTAGTAACCT 17  
DB 79210 AGCGCGTAGTAACCT 79194



RESULT 4  
AA199683\_14/c  
Continuation (15 of 44) of AA199683 from base 1400001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683  
WP Fragment Name Begin End  
WP AA199683\_00 1 110000  
WP AA199683\_01 100001 210000  
WP AA199683\_02 200001 310000  
WP AA199683\_03 300001 410000  
WP AA199683\_04 400001 510000  
WP AA199683\_05 500001 610000  
WP AA199683\_06 600001 710000  
WP AA199683\_07 700001 810000  
WP AA199683\_08 800001 910000  
WP AA199683\_09 900001 1010000  
WP AA199683\_10 1000001 1110000  
WP AA199683\_11 1100001 1210000  
WP AA199683\_12 1200001 1310000  
WP AA199683\_13 1300001 1410000  
WP AA199683\_14 1400001 1510000  
WP AA199683\_15 1500001 1610000  
WP AA199683\_16 1600001 1710000  
WP AA199683\_17 1700001 1810000  
WP AA199683\_18 1800001 1910000  
WP AA199683\_19 1900001 2010000  
WP AA199683\_20 2000001 2110000  
WP AA199683\_21 2100001 2210000  
WP AA199683\_22 2200001 2310000  
WP AA199683\_23 2300001 2410000  
WP AA199683\_24 2400001 2510000  
WP AA199683\_25 2500001 2610000  
WP AA199683\_26 2600001 2710000  
WP AA199683\_27 2700001 2810000  
WP AA199683\_28 2800001 2910000  
WP AA199683\_29 2900001 3010000  
WP AA199683\_30 3000001 3110000  
WP AA199683\_31 3100001 3210000  
WP AA199683\_32 3200001 3310000  
WP AA199683\_33 3300001 3410000  
WP AA199683\_34 3400001 3510000  
WP AA199683\_35 3500001 3610000  
WP AA199683\_36 3600001 3710000  
WP AA199683\_37 3700001 3810000  
WP AA199683\_38 3800001 3910000  
WP AA199683\_39 3900001 4010000  
WP AA199683\_40 4000001 4110000  
WP AA199683\_41 4100001 4210000  
WP AA199683\_42 4200001 4310000  
WP AA199683\_43 4300001 4403765

Query Match 100.0%; Score 17; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGCGTAGTACT 17  
Db 78734 AGCCGCGTAGTACT 78718

RESULT 5  
AAAX21192  
ID AAAX21192 standard; DNA; 505 BP.  
XX  
XX  
AC AAAX21192;  
XX  
XX  
DT 05-MAY-1999 (first entry)  
XX  
XX  
DE Polynucleotide sequence from the genome of *Treponema pallidum*.  
XX  
XX  
KW *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;  
KV enzyme production; ds.  
XX  
OS *Treponema pallidum*.

XX  
PN W09859034-A2.  
XX  
XX 30-DEC-1998.  
PD  
XX  
PF 23-JUN-1998; 98MO-US013041.  
XX  
XX  
PR 24-JUN-1997; 97US-0050667P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Frazer CM;  
XX  
XX WPI; 1999-081273/07.  
XX  
XX  
PT New isolated *Treponema pallidum* nucleic acids - used to develop products  
PT for the detection, diagnosis, characterisation, prevention and therapy of  
XX T. *pallidum* infections, particularly syphilis.  
XX  
PS Claim 1; Page 1116; 1150pp; English.  
XX  
XX  
CC AAX20500-21243 represent polynucleotide sequences from the genome of  
CC *Treponema pallidum*. The sequences can be used for detection, diagnosis,  
CC characterisation, prevention and therapy for T. *pallidum* infections,  
CC particularly syphilis. They can also be used for detecting diseases  
CC related to *Borrelia* infections in animals, and for the production of  
CC biosynthetic products such as enzymes

SQ Sequence 505 BP; 87 A; 127 C; 153 G; 132 T; 0 U; 6 Other;  
Query Match 84.7%; Score 14.4; DB 2; Length 505;  
Best Local Similarity 93.8%; Pred. No. 3e+02; 1;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GCCGCGTAGTACT 17  
Db 429 GCCGCGTAGTACT 444

RESULT 6  
ACA54363/c  
ID ACA54363 standard; DNA; 735 BP.  
XX  
XX  
AC ACA54363;  
XX  
XX  
DT 19-JUN-2003 (first entry)  
XX  
XX  
DE Prokaryotic essential gene #36020.  
XX  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
XX  
OS *Yersinia pestis*.  
XX  
XX  
PN W0200277183-A2.  
PD  
XX 03-OCT-2002.  
XX  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342922P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX  
XX P-PSDB; ABUS0493.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 42233; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway; (8)  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the cell compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 735 BP; 180 A; 141 C; 214 G; 200 T; 0 U; 0 Other;  
XX  
Query Match 84.7%; Score 14.4; DB 8; Length 735;  
Best Local Similarity 93.8%; Pred. No. 3.1e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GCGCGTAGGTAACT 17  
DB 458 GCGCGTAGGTAACT 443  
XX  
RESULT 7  
AAS93164/c  
ID AAS93164 standard; cDNA; 2967 BP.  
XX  
AC AAS93164;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #28968.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
XX  
XX 31-MAR-2000; 2000US-00540217.  
XX  
XX 23-AUG-2000; 2000US-00649167.  
XX  
XX

XX  
PA (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX MPI; 2001-639362/73.  
XX  
DR P-PSDB; ABG28977.  
XX  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
PS Claim 1; SEQ ID NO 28968; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal  
XX activity of (II) or to treat disease states involving (II). (II) is  
XX useful for generating antibodies against it, detecting or quantitating a  
XX polypeptide in tissue, as molecular weight markers and as a food  
XX supplement. (II) and its binding partners are useful in medical imaging  
XX of sites expressing (II). (I) and (II) are useful for treating disorders  
XX involving aberrant protein expression or biological activity. The  
XX polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS94197-AAS94564 represent novel human diagnostic  
XX coding sequences of the invention. Note: The sequence data for this  
XX patent did not appear in the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2967 BP; 689 A; 782 C; 772 G; 724 T; 0 U; 0 Other;  
XX  
Query Match 84.7%; Score 14.4; DB 5; Length 2967;  
Best Local Similarity 93.8%; Pred. No. 3.3e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GCGCGTAGGTAACT 17  
DB 2267 GCGCGTAGGTAACT 2252  
XX  
RESULT 8  
ACT21756  
ID ACT21756 standard; DNA; 237 BP.  
XX  
AC ACT21756;  
XX  
DT 27-OCT-2003 (revised)  
XX  
DT 17-OCT-2003 (first entry)  
XX  
DE DNA clone originating in barley containing SNP encoding sequence #11747.  
XX  
XX Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
XX gene; ss.  
XX  
XX Hordeum vulgare; ssp. spontaneum.  
XX  
XX WO2003057877-A1.  
XX  
XX 17-JUL-2003.  
XX  
XX 16-DEC-2002; 2002WO-IB005403.  
XX  
XX 20-DEC-2001; 2001JP-00387059.  
XX  
XX 20-DEC-2001; 2001JP-00387131.  
XX  
XX 20-DEC-2001; 2001JP-00403299.  
XX  
XX

PR 20-DEC-2001; 2001JP-00403300.  
PR 27-SEP-2002; 2002JP-00327515.  
XX  
PA (UYN1-) UNIV JAPAN OKAYAMA.  
XX  
PI Sato K, Takeda K, Kohara Y;  
XX  
DR WPI; 2003-587127/55.  
XX  
PT Single nucleotide polymorphism sites in barley varieties and DNA  
PT sequences containing them for analysis and identification of barley  
PT varieties and production of barley transformants with desired  
PT characteristics.  
XX  
PS Disclosure; SEQ ID XX; 284bp; Japanese.  
XX  
CC The present invention relates to oligonucleotide clones originating in  
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
CC varieties, identification of particular varieties and genotype-phenotype  
CC analysis, isolation of specific genes and creation of new varieties by  
CC transformation of barley varieties with them and production of new barley  
CC varieties with desired properties. The present sequence represents an  
CC oligonucleotide clone DNA sequence featured in the specification. The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 237 BP; 33 A; 80 C; 64 G; 59 T; 0 U; 1 Other;  
XX  
Query Match 82.4%; Score 14; DB 9; Length 237;  
Best Local Similarity 100.0%; Pred. No. 5e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 2 GCCGCGTAGTAAC 15  
DB 88 GCCGCGTAGTAAC 101  
XX  
RESULT 9  
ACL21751  
ID ACL21751 standard; DNA; 317 BP.  
XX  
AC ACL21751;  
XX  
DT 27-OCT-2003 (revised)  
DT 17-OCT-2003 (first entry)  
XX  
DE DNA clone originating in barley containing SNP encoding sequence #11742.  
XX  
KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
KM gene; ss.  
XX  
OS Hordeum vulgare; ssp. spontaneum.  
XX  
PN WO2003057877-A1.  
PD 17-JUL-2003.  
XX  
PF 16-DEC-2002; 2002WO-1B005403.  
XX  
PR 20-DEC-2001; 2001JP-00387059.  
PR 20-DEC-2001; 2001JP-00387131.  
PR 20-DEC-2001; 2001JP-00403299.  
PR 20-DEC-2001; 2001JP-00403300.  
PR 27-SEP-2002; 2002JP-00327515.  
XX  
PA (UYN1-) UNIV JAPAN OKAYAMA.  
XX  
PI Sato K, Takeda K, Kohara Y;  
XX  
DR WPI; 2003-587127/55.

XX  
PT Single nucleotide polymorphism sites in barley varieties and DNA  
PT sequences containing them for analysis and identification of barley  
PT varieties and production of barley transformants with desired  
PT characteristics.  
XX  
PS Disclosure; SEQ ID XX; 284bp; Japanese.  
XX  
CC The present invention relates to oligonucleotide clones originating in  
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
CC varieties, identification of particular varieties and genotype-phenotype  
CC analysis, isolation of specific genes and creation of new varieties by  
CC transformation of barley varieties with them and production of new barley  
CC varieties with desired properties. The present sequence represents an  
CC oligonucleotide clone DNA sequence featured in the specification. The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 317 BP; 60 A; 95 C; 91 G; 71 T; 0 U; 0 Other;  
XX  
Query Match 82.4%; Score 14; DB 9; Length 317;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 2 GCCGCGTAGTAAC 15  
DB 268 GCCGCGTAGTAAC 281  
XX  
RESULT 10  
ADR91508/c  
ID ADR91508 standard; DNA; 360 BP.  
XX  
AC ADR91508;  
XX  
DT 16-DEC-2004 (first entry)  
XX  
DE Novel S. pneumoniae DNA sequence, SEQ ID 143.  
XX  
KM Meningitis; bacteraemia; pneumonia; otitis media; ds;  
KM bacterial infection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US6800744-B1.  
PD 05-OCT-2004.  
XX  
PF 30-JUN-1998; 98US-00107433.  
XX  
PR 02-JUL-1997; 97US-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2004-697205/68.  
DR P-PSDB; ADR94111.  
XX  
PT New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 143; 151bp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,

CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridizable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC *Streptococcus pneumoniae* e.g. pneumonia, bacteremia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC *pneumoniae* nucleic acid sequences. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC [seqdata.uspto.gov/sequence.html?docID=6800744B1](http://seqdata.uspto.gov/sequence.html?docID=6800744B1).  
CC  
CC Sequence 360 BP; 80 A; 98 C; 59 G; 123 T; 0 U; 0 Other;  
SQ  
Query Match 82.4%; Score 14; DB 13; Length 360;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02; Mismatches 0; Gaps 0;  
Matches 14; Conservative 0; Indels 0; Gaps 0;  
QY 4 CCGGTAGTAACT 17  
DB 49 CCGGTAGTAACT 36  
RESULT 11  
ACL21733  
XX ACL21733 standard; DNA; 394 BP.  
XX  
XX ACL21733;  
XX  
XX 27-OCT-2003 (revised)  
XX DT 17-OCT-2003 (first entry)  
XX  
XX DNA clone originating in barley containing SNP encoding sequence #11724.  
XX  
XX Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
XX gene; ss.  
XX  
XX Hordeum vulgare; var. (cul.Haruna Nijo).  
XX OS  
XX PN WO2003057877-A1.  
XX PD 17-JUL-2003.  
XX  
XX 16-DEC-2002; 2002WO-IB005403.  
XX PF  
XX PR 20-DEC-2001; 2001JP-00387059.  
XX PR 20-DEC-2001; 2001JP-00387131.  
XX PR 20-DEC-2001; 2001JP-00403299.  
XX PR 20-DEC-2001; 2001JP-00403300.  
XX PR 27-SRP-2002; 2002JP-00327515.  
XX  
XX (UYNT-) UNIV JAPAN OKAYAMA.  
XX PA  
XX PI Sato K, Takeda K, Kohara Y;  
XX  
XX WPI; 2003-587127/55.  
XX  
XX Single nucleotide polymorphism sites in barley varieties and DNA  
XX PT sequences containing them for analysis and identification of barley  
XX PT varieties and production of barley transformants with desired  
XX PT characteristics.  
XX  
XX PS Disclosure; SEQ ID XX; 284bp; Japanese.  
XX  
XX The present invention relates to oligonucleotide clones originating in

CC barley (*Hordeum vulgare*) which contain single nucleotide polymorphisms  
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
CC varieties, identification of particular varieties and genotype-phenotype  
CC analysis, isolation of specific genes and creation of new varieties by  
CC transformation of barley varieties with them and production of new barley  
CC varieties with desired properties. The present sequence represents an  
CC oligonucleotide clone DNA sequence featured in the specification. The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published-pct-sequences](http://ftp.wipo.int/pub/published-pct-sequences). (updated on 27-OCT-2003 to  
CC standardise OS field)  
CC  
XX Sequence 394 BP; 70 A; 121 C; 107 G; 96 T; 0 U; 0 Other;  
SQ  
Query Match 82.4%; Score 14; DB 9; Length 394;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02; Mismatches 0; Gaps 0;  
Matches 14; Conservative 0; Indels 0; Gaps 0;  
QY 2 GCCGCTAGTAAAC 15  
DB 260 GCCGCTAGTAAAC 273  
RESULT 12  
ACL21749  
XX ACL21749 standard; DNA; 553 BP.  
XX  
XX ACL21749;  
XX  
XX 27-OCT-2003 (revised)  
XX DT 17-OCT-2003 (first entry)  
XX  
XX DNA clone originating in barley containing SNP encoding sequence #11740.  
XX  
XX Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
XX gene; ss.  
XX  
XX Hordeum vulgare; ssp. spontaneum.  
XX KM  
XX PN WO2003057877-A1.  
XX PD 17-JUL-2003.  
XX  
XX 16-DEC-2002; 2002WO-IB005403.  
XX PF  
XX PR 20-DEC-2001; 2001JP-00387059.  
XX PR 20-DEC-2001; 2001JP-00387131.  
XX PR 20-DEC-2001; 2001JP-00403299.  
XX PR 20-DEC-2001; 2001JP-00403300.  
XX PR 27-SRP-2002; 2002JP-00327515.  
XX  
XX (UYNT-) UNIV JAPAN OKAYAMA.  
XX PA  
XX PI Sato K, Takeda K, Kohara Y;  
XX  
XX WPI; 2003-587127/55.  
XX  
XX Single nucleotide polymorphism sites in barley varieties and DNA  
XX PT sequences containing them for analysis and identification of barley  
XX PT varieties and production of barley transformants with desired  
XX PT characteristics.  
XX  
XX PS Disclosure; SEQ ID XX; 284bp; Japanese.  
XX  
XX The present invention relates to oligonucleotide clones originating in  
XX CC barley (*Hordeum vulgare*) which contain single nucleotide polymorphisms  
XX CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
XX CC varieties, identification of particular varieties and genotype-phenotype  
XX CC analysis, isolation of specific genes and creation of new varieties by  
XX CC transformation of barley varieties with them and production of new barley  
XX CC varieties with desired properties. The present sequence represents an  
XX CC oligonucleotide clone DNA sequence featured in the specification. The  
XX CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
CC standardise OS field)

XX Sequence 553 BP; 134 A; 165 C; 136 G; 114 T; 0 U; 4 Other;

Query Match 82.4%; Score 14; DB 9; Length 553;  
Best Local Similarity 100.0%; Pred. No. 5.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCGCGGTAGGTAC 15  
DB 513 GCGCGGTAGGTAC 526

RESULT 13  
ACLT1732  
ID ACLT1732 standard; DNA; 561 BP.

XX ACLT1732;

DT 27-OCT-2003 (revised)  
DT 17-OCT-2003 (first entry)

DE DNA clone originating in barley containing SNP encoding sequence #11723.

KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
KM gene; ss.

OS Hordeum vulgare; var. (cul.Haruna Nijo).

PN WO2003057877-A1.

XX 17-JUL-2003.

PD 16-DEC-2002; 2002MO-IB005403.

XX 20-DEC-2001; 2001JP-00387059.

PR 20-DEC-2001; 2001JP-00387131.

PR 20-DEC-2001; 2001JP-00403299.

PR 27-SEP-2002; 2002JP-00327515.

XX (UYNI-) UNIV JAPAN OKAYAMA.

XX Sato K, Takeda K, Kohara Y;

XX WPI; 2003-587127/55.

XX Single nucleotide polymorphism sites in barley varieties and DNA  
PT sequences containing them for analysis and identification of barley  
PT varieties and production of barley transformants with desired  
PT characteristics.

XX Disclosure; SEQ ID XX; 284pp; Japanese.

XX The present invention relates to oligonucleotide clones originating in  
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
CC varieties, identification of particular varieties and genotype-phenotype  
CC analysis, isolation of specific genes and creation of new varieties by  
CC transformation of barley varieties with them and production of new barley  
CC varieties with desired properties. The present sequence represents an  
CC oligonucleotide clone DNA sequence featured in the specification. The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
CC standardise OS field)

XX Sequence 561 BP; 111 A; 166 C; 148 G; 136 T; 0 U; 0 Other;

Query Match 82.4%; Score 14; DB 9; Length 561;  
Best Local Similarity 100.0%; Pred. No. 5.2e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCGCGGTAGGTAC 15

DB 415 GCGCGGTAGGTAC 428

RESULT 14  
ACLT1772  
ID ACLT1772 standard; DNA; 582 BP.

XX ACLT1772;

DT 27-OCT-2003 (revised)  
DT 17-OCT-2003 (first entry)

DE DNA clone originating in barley containing SNP encoding sequence #11763.

KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
KM gene; ss.

OS Hordeum vulgare; ssp. spontaneum.

PN WO2003057877-A1.

XX 17-JUL-2003.

PD 16-DEC-2002; 2002MO-IB005403.

XX 20-DEC-2001; 2001JP-00387059.

PR 20-DEC-2001; 2001JP-00387131.

PR 20-DEC-2001; 2001JP-00403299.

PR 27-SEP-2002; 2002JP-00327515.

XX (UYNI-) UNIV JAPAN OKAYAMA.

XX Sato K, Takeda K, Kohara Y;

XX WPI; 2003-587127/55.

XX Single nucleotide polymorphism sites in barley varieties and DNA  
PT sequences containing them for analysis and identification of barley  
PT varieties and production of barley transformants with desired  
PT characteristics.

XX Disclosure; SEQ ID XX; 284pp; Japanese.

XX The present invention relates to oligonucleotide clones originating in  
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
CC varieties, identification of particular varieties and genotype-phenotype  
CC analysis, isolation of specific genes and creation of new varieties by  
CC transformation of barley varieties with them and production of new barley  
CC varieties with desired properties. The present sequence represents an  
CC oligonucleotide clone DNA sequence featured in the specification. The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
CC standardise OS field)

XX Sequence 582 BP; 139 A; 174 C; 140 G; 124 T; 0 U; 5 Other;

Query Match 82.4%; Score 14; DB 9; Length 582;  
Best Local Similarity 100.0%; Pred. No. 5.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCGCGGTAGGTAC 15  
DB 521 GCGCGGTAGGTAC 534

RESULT 15

AC121729  
 ID ACL21729 standard; DNA; 591 BP.  
 XX  
 AC ACL21729;  
 XX  
 DT 27-OCT-2003 (revised)  
 DT 17-OCT-2003 (first entry)  
 XX  
 DE DNA clone originating in barley containing SNP encoding sequence #11720.  
 XX  
 KM Barley, single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
 KM gene; ss.  
 XX  
 OS Hordeum vulgare; var. (cul.Haruna Nijo).  
 XX  
 PN WO2003057877-A1.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 16-DEC-2002; 2002WO-1B005403.  
 XX  
 PR 20-DEC-2001; 2001JP-00387059.  
 PR 20-DEC-2001; 2001JP-00387131.  
 PR 20-DEC-2001; 2001JP-00403299.  
 PR 20-DEC-2001; 2001JP-00403300.  
 PR 27-SEP-2002; 2002JP-00327515.  
 XX  
 PA (UYN1-) UNTV JAPAN OKAYAMA.  
 XX  
 PI Sato K, Takeda K, Kohara Y;  
 XX  
 DR MPI; 2003-587127/55.  
 XX  
 PT Single nucleotide polymorphism sites in barley varieties and DNA  
 PT sequences containing them for analysis and identification of barley  
 PT varieties and production of barley transformants with desired  
 PT characteristics.  
 XX  
 PS Disclosure; SEQ ID XX; 284bp; Japanese.  
 XX  
 CC The present invention relates to oligonucleotide clones originating in  
 CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
 CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
 CC varieties, identification of particular varieties and genotype-phenotype  
 CC analysis, isolation of specific genes and creation of new varieties by  
 CC transformation of barley varieties with them and production of new barley  
 CC varieties with desired properties. The present sequence represents an  
 CC oligonucleotide clone DNA sequence featured in the specification. The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 591 BP; 103 A; 162 C; 139 G; 186 T; 0 U; 1 Other;  
 Query Match 82.4%; Score 14; DB 9; Length 591;  
 Best Local Similarity 100.0%; Pred.No. 5.2e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GCCGCGTAGGTAAC 15  
 |||||  
 Db 60 GCCGCGTAGGTAAC 73

Search completed: November 7, 2005, 05:40:40  
 Job time : 203.043 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 20:56:42 ; Search time 423.915 Seconds  
(without alignments)  
331.640 Million cell updates/sec

Title: US-10-777-131A-5

Perfect score: 17  
Sequence: 1 agccgcgtagtaacct 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:\*  
23: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq:\*  
26: /cgn2\_6/ptodata/2/pubpna/US11A\_NEW\_PUB.seq:\*  
27: /cgn2\_6/ptodata/2/pubpna/US66\_NEW\_PUB.seq:\*  
28: /cgn2\_6/ptodata/2/pubpna/US66\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	100.0	17	US-10-216-817-5	Sequence 5, Appli
2	17	100.0	17	US-10-777-131A-5	Sequence 5, Appli
3	17	100.0	2488	US-10-216-817-17	Sequence 17, Appli
4	17	100.0	2488	US-10-777-131A-17	Sequence 17, Appli
5	15	88.2	526	US-10-425-115-180069	Sequence 180069,

C	6	14.4	84.7	473	13	US-09-925-065A-480693	Sequence 480693,
C	7	14.4	84.7	735	18	US-10-282-122A-42233	Sequence 42233, A
C	8	14.4	2967	14	US-10-450-763-28968	Sequence 28968, A	
9	14.4	84.7	3186778	24	US-10-027-632-174961	Sequence 174961,	
10	14.4	84.7	3186778	18	US-10-027-632-174961	Sequence 174961,	
11	14	82.4	360	24	US-10-617-320-143	Sequence 143, App	
12	14	82.4	689	14	US-10-027-632-240523	Sequence 240523,	
13	14	82.4	689	18	US-10-027-632-240523	Sequence 240523,	
14	14	82.4	1334	9	US-09-737-190A-2	Sequence 2, Appli	
15	14	82.4	1335	17	US-10-029-397A-31	Sequence 31, Appli	
16	14	82.4	1385	14	US-10-004-717-30	Sequence 9, Appli	
17	14	82.4	1385	18	US-10-425-259-17	Sequence 17, Appli	
18	14	82.4	1385	21	US-10-860-373-30	Sequence 30, Appli	
19	14	82.4	1385	21	US-10-860-724-30	Sequence 30, Appli	
20	14	82.4	1515	9	US-09-726-774-9	Sequence 9, Appli	
21	14	82.4	1515	17	US-10-029-397A-30	Sequence 30, Appli	
22	14	82.4	1515	20	US-10-719-633-9	Sequence 9, Appli	
23	14	82.4	1522	18	US-10-461-990-1	Sequence 1, Appli	
24	14	82.4	1551	17	US-10-029-397A-29	Sequence 29, Appli	
25	14	82.4	6109	10	US-09-795-061-1	Sequence 1, Appli	
26	14	82.4	21338	8	US-08-961-527-20	Sequence 20, Appli	
27	14	82.4	21338	18	US-10-158-84A-20	Sequence 20, Appli	
28	14	82.4	2162598	22	US-10-472-928-4979	Sequence 4979, Ap	
C	29	14	82.4	2162598	22	US-10-472-928-4979	Sequence 4979, Ap
C	30	13.8	81.2	201	20	US-10-741-601-21465	Sequence 21465, A
C	31	13.8	81.2	201	22	US-10-741-601-60754	Sequence 60754, A
C	32	13.8	81.2	258	21	US-10-425-115-161222	Sequence 161222,
C	33	13.8	81.2	268	21	US-10-425-115-183778	Sequence 183778,
C	34	13.8	81.2	330	19	US-10-424-599-14635	Sequence 14635, A
C	35	13.8	81.2	385	21	US-10-425-115-8899	Sequence 8899, Ap
C	43	81.2	473	19	US-10-424-599-137820	Sequence 137820,	
C	44	81.2	519	21	US-10-425-115-170941	Sequence 170941,	
C	45	13.8	81.2	532	14	US-10-027-632-280336	Sequence 280336,
						Sequence 290337,	

#### ALIGNMENTS

RESULT 1  
US-10-216-817-5  
; Sequence 5, Application US/10216817  
; Publication No. US20030129619A1  
GENERAL INFORMATION:  
APPLICANT: GICOURT, BRIGITTE  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
FILE REFERENCE: 03495.0233-00000  
CURRENT APPLICATION NUMBER: US/10/216,817  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 60/311,824  
PRIOR FILING DATE: 2001-08-14  
PRIOR APPLICATION NUMBER: 60/313,523  
PRIOR FILING DATE: 2001-08-21  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 17  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURES:  
OTHER INFORMATION: Description of Artificial Sequence: Primer  
US-10-216-817-5  
Query Match 100.0%; Score 17; DB 16; Length 17;  
Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17  
|||||  
Db 1 AGCCGCGTAGTAACCT 17

RESULT 2  
US-10-777-131A-5  
; Sequence 5, Application US/10777131A  
; Publication No. US20050026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT APPLICATION NUMBER: US/10/777,131A  
; CURRENT FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 5  
; LENGTH: 17  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer  
US-10-777-131A-5

Query Match 100.0%; Score 17; DB 22; Length 17;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17  
|||||  
Db 1 AGCCGCGTAGTAACCT 17

RESULT 3  
US-10-216-817-17  
; Sequence 17, Application US/10216817  
; Publication No. US20030129619A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; FILE REFERENCE: 03495.0233-00000  
; CURRENT APPLICATION NUMBER: US/10/216,817  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 2488  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-216-817-17

Query Match 100.0%; Score 17; DB 16; Length 2488;  
Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17

Db 407 AGCCGCGTAGTAACCT 423  
|||||

RESULT 4  
US-10-777-131A-17  
; Sequence 17, Application US/10777131A  
; Publication No. US20050026216A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
; FILE REFERENCE: 02356.0090-00000  
; CURRENT APPLICATION NUMBER: US/10/777,131A  
; CURRENT FILING DATE: 2004-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/09679  
; PRIOR FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/311,824  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,523  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 17  
; LENGTH: 2488  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-777-131A-17

Query Match 100.0%; Score 17; DB 22; Length 2488;  
Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17  
|||||  
Db 407 AGCCGCGTAGTAACCT 423

RESULT 5  
US-10-425-115-180069/c  
; Sequence 180069, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(5322)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 180069  
; LENGTH: 526  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(526)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_95802C.1  
US-10-425-115-180069

Query Match 88.2%; Score 15; DB 21; Length 526;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTAACCT 16  
|||||  
Db 127 GCCGCGTAGTAACCT 113



```
RESULT 6
US-09-925-065A-480693/C
; Sequence 480693, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 480693
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-480693

Query Match
Best Local Similarity 84.7%; Score 14.4; DB 13; Length 473;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCCGCGTAGTAACC 16
Db 210 ACCCGCGCGAGTAACC 195

RESULT 7
US-10-282-122A-42233/C
; Sequence 42233, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42233
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Yersinia pestis
US-10-282-122A-42233

Query Match
Best Local Similarity 84.7%; Score 14.4; DB 18; Length 735;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTAACCT 17
Db 458 GCCGCGTAGTAACCT 443

RESULT 8
US-10-450-763-28968/C
; Sequence 28968, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: HyBeq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 28968
; LENGTH: 2967
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (2488)..(2769)
; OTHER INFORMATION: 65% homologous to Salmonella typhimurium Lr2 92% identity
; OTHER INFORMATION: with E. coli lipopolysaccharide biosynthesis protein (WECE)
; OTHER INFORMATION: (SP:P27833); contains similarity to Pfam family PF01041,
; OTHER INFORMATION: accession number AF233324, Smith-Waterman Score=293.
US-10-450-763-28968

Query Match
Best Local Similarity 84.7%; Score 14.4; DB 24; Length 2967;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTAACCT 17
Db 2267 GCCGCGTAGTAACCT 2252

RESULT 9
US-10-027-632-174961
; Sequence 174961, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

TITLE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 174961  
LENGTH: 3186778  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(3186778)  
US-10-027-632-174961

Query Match 84.7%; Score 14.4; DB 14; Length 3186778;  
Best Local Similarity 93.8%; Pred. No. 2.6e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCGCGTAGTAACCT 17  
DB 553211 GCGCGTAGGAACCT 553226

RESULT 10  
US-10-027-632-174961  
Sequence 174961, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 174961  
LENGTH: 3186778  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(3186778)

OTHER INFORMATION: n = A,T,C or G  
US-10-027-632-174961

Query Match 84.7%; Score 14.4; DB 18; Length 3186778;  
Best Local Similarity 93.8%; Pred. No. 2.6e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCGCGTAGTAACCT 17  
DB 553211 GCGCGTAGGAACCT 553226

RESULT 11  
US-10-617-320-143/C  
Sequence 143, Application US/10617320  
Publication No. US20050136404A1  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
NUMBER OF SEQUENCES: 5206  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/617,320  
FILING DATE: 10-Jul-2003  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Atinidello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 143:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...360  
SEQUENCE DESCRIPTION: SEQ ID NO: 143:  
US-10-617-320-143  
Query Match 82.4%; Score 14; DB 24; Length 360;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 49 CGCGTAGTAGTAACT 36

RESULT 12  
US-10-027-632-240523

Sequence 240523, Application US/10027632  
Publication No. US2002019837A1  
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 240523

LENGTH: 689

TYPE: DNA

ORGANISM: Human

US-10-027-632-240523

Query Match

Best Local Similarity 82.4%; Score 14; DB 14; Length 689;

Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAGTAACT 16

DB 422 AGCCGCGTAGTAGTAACT 437

RESULT 13

US-10-027-632-240523

Sequence 240523, Application US/10027632

Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 240523

LENGTH: 689

TYPE: DNA

ORGANISM: Human

US-10-027-632-240523

Query Match

Best Local Similarity 82.4%; Score 14; DB 14; Length 689;

Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAGTAACT 16

DB 422 AGCCGCGTAGTAGTAACT 437

RESULT 14

US-09-737-190A-2

Sequence 2, Application US/09737190A

Patent No. US20020102545A1

GENERAL INFORMATION:

APPLICANT: Shibuya, Tetsuo

TITLE OF INVENTION: A Method for Changing a Target Array, a Method for Analyzing

STRUCTURE, AND AN APPARATUS, A STORAGE MEDIUM AND A

TRANSMISSION MEDIUM THEREFOR

FILE REFERENCE: JP91990270US1 (14043)

CURRENT APPLICATION NUMBER: US/09/737,190A

PRIOR FILING DATE: 2000-12-14

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 1334

TYPE: DNA

ORGANISM: Streptococcus anginosus

FEATURE:

NAME/KEY: UNSURE

LOCATION: (412)

OTHER INFORMATION: n at position 412 is uncertain

NAME/KEY: UNSURE

LOCATION: (449)

OTHER INFORMATION: n at position 449 is uncertain

NAME/KEY: UNSURE

LOCATION: (701)

OTHER INFORMATION: n at position 701 is uncertain

NAME/KEY: UNSURE

LOCATION: (855)

OTHER INFORMATION: n at position 855 is uncertain

NAME/KEY: UNSURE

LOCATION: (869)

OTHER INFORMATION: n at position 869 is uncertain

NAME/KEY: UNSURE

LOCATION: (909)

OTHER INFORMATION: n at position 909 is uncertain

NAME/KEY: UNSURE

LOCATION: (1018)

OTHER INFORMATION: n at position 1018 is uncertain

NAME/KEY: UNSURE

LOCATION: (1070)

OTHER INFORMATION: n at position 1070 is uncertain

NAME/KEY: UNSURE

LOCATION: (1121)

OTHER INFORMATION: n at position 1121 is uncertain

NAME/KEY: UNSURE

LOCATION: (1228)

OTHER INFORMATION: n at position 1228 is uncertain

NAME/KEY: UNSURE

LOCATION: (1304)

OTHER INFORMATION: n at position 1304 is uncertain

US-09-737-190A-2

Query Match

Best Local Similarity 100.0%; Score 14; DB 9; Length 1334;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 CGCGTAGGTAACT 17  
 |||||  
 Db 15 CGCGTAGGTAACT 28

RESULT 15  
 US-10-029-397A-31  
 ; Sequence 31, Application US/10029397A  
 ; Publication No. US20030175709A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MURPHY, GEORGE L.  
 ; APPLICANT: WHITLEY, J. PENN  
 ; TITLE OF INVENTION: METHOD AND SYSTEM FOR DEPLETING RNA POPULATIONS  
 ; FILE REFERENCE: AMB1:076US  
 ; CURRENT APPLICATION NUMBER: US/10/029,397A  
 ; CURRENT FILING DATE: 2002-03-19  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 31  
 ; LENGTH: 1335  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pyogenes  
 US-10-029-397A-31

Query Match 82.4%; Score 14; DB 17; Length 1335;  
 Best Local Similarity 100.0%; Pred.No.4.5e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 CGCGTAGGTAACT 17  
 |||||  
 Db 15 CGCGTAGGTAACT 28

Search completed: November 7, 2005, 05:21:31  
 Job time : 440.915 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 1700.72 Seconds  
(without alignments)  
380.481 Million cell updates/sec

Title: US-10-777-131A-5

Perfect score: 17  
Sequence: 1 agccgcgtagtaacct 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3423544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_esc1:\*  
2: gb\_esc2:\*  
3: gb\_hic:\*  
4: gb\_esc3:\*  
5: gb\_esc4:\*  
6: gb\_esc5:\*  
7: gb\_esc6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	94.1	588	6	CD811742 BM10.001L
2	15.4	90.6	235	7	CV101208 UMC-PCR7b
3	15.4	90.6	411	4	BM539228 h206f02.9
4	15.4	90.6	528	6	CA217548 SCRUD113
5	15.4	90.6	533	7	CV224495 CS_hyp_19
6	15.4	90.6	667	7	CV224494 CS_hyp_19
7	15.4	90.6	700	5	BO294905 WHE2855 G
8	15.4	90.6	768	6	CL653433 PRI0118C
9	15	88.2	509	6	CD445870 E101T0204
10	15	88.2	724	9	CNS0607P T3 end of
11	15	88.2	821	8	BZ797610 PUFRC69TD
12	15	88.2	927	9	CG063514 PUKARS9TD
13	15	88.2	1107	2	BF672870 602152718
14	15	88.2	398	7	CK914339 PJfmcfc_0
15	14.4	84.7	419	1	AJ828384 AJ828384
16	14.4	84.7	472	7	CF604790 RADIC01_0
17	14.4	84.7	480	1	AJ282503 AA3A-AY-
18	14.4	84.7	487	5	BQ499259 EST08484
19	14.4	84.7	520	5	BP743576 BP743576
20	14.4	84.7	535	5	BO503246 EST12476
21	14.4	84.7	546	7	CF604144 BACCA01.0
22	14.4	84.7	558	8	AZ438618 1M0228K14
23	14.4	84.7	578	5	BX602283 BX602283
24	14.4	84.7	596	4	BM635084 BM635084

25	14.4	84.7	608	7	CK708350 ZF201-P00
26	14.4	84.7	614	4	BM651965 170006873
27	14.4	84.7	622	3	CNS09XEG BX062516 Single re
28	14.4	84.7	628	5	BM170980 BM170980
29	14.4	84.7	639	4	BM613376 170006591
30	14.4	84.7	647	4	BM602408 170006870
31	14.4	84.7	649	4	BM589600 170006684
32	14.4	84.7	651	5	BX603222 BX603222
33	14.4	84.7	657	1	AY896818 AY896818
34	14.4	84.7	660	8	AQ952838 Sheared D
35	14.4	84.7	665	4	BM612076 170006871
36	14.4	84.7	666	2	BE306697 601104091
37	14.4	84.7	668	5	BM472135 BM472135
38	14.4	84.7	670	4	BM621607 170006874
39	14.4	84.7	681	2	BF124082 601760175
40	14.4	84.7	685	4	BM632505 170006875
41	14.4	84.7	690	5	BM015565 BM015565
42	14.4	84.7	706	4	BM652288 170006873
43	14.4	84.7	709	5	BM013708 BM013708
44	14.4	84.7	714	4	BM607359 BM607359
45	14.4	84.7	714	8	BZ628070 INS7112.B

#### ALIGNMENTS

RESULT 1  
CD811742  
LOCUS  
DEFINITION  
BNI0.00106F011207 BNI0 Braasica napus cDNA clone BNI0001106, mRNA  
sequence.  
ACCESSION  
CD811742  
VERSION  
CD811742.1 GI:32493582  
KEYWORDS  
EST.  
SOURCE  
Braasica napus (rape)  
ORGANISM  
Braasica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE  
1 (bases 1 to 588)  
AUTHORS  
Genopiante.  
TITLE  
Unpublished (2003)  
JOURNAL  
CONTACT  
Genopiante  
COMMENT  
Genopiante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genopiante' (<http://www.genopiante.com>  
and <http://genopiante-info.inbio.gen.fr>).  
Location/Qualifiers

FEATURES  
source  
1..588  
/organism="Braasica napus"  
/mol\_type="mRNA"  
/cultivar="Jec Neuf"  
/db\_xref="taxon:3708"  
/clone="BNI0001106"  
/issue\_type="seed"  
/clone\_1ib="BNI0"

#### ORIGIN

Query Match 94.1%; Score 16; DB 6; Length 588;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTAACT 17  
|||||  
DB 526 GCCGCGTAGTAACT 541

RESULT 2  
CV101208

LOCUS CV101208 235 bp mRNA linear EST 09-SEP-2004  
 DEFINITION UMC-pcrtbcl1-006-f01 Blastocyst after cryopreservation pcrtbcl Sus  
 ACCESSION CV101208  
 VERSION CV101208.1 GI:51583372  
 KEYWORDS EST.  
 SOURCE Sus scrofa (pig)  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 1 (bases 1 to 235)  
 Woods, R., Whitworth, K., Springer, G.K., Forrester, L.J.,  
 Spollen, W.G., Riese, J., Bivens, N., Murphy, C.N., Mathialigan, N.,  
 Agca, Y., Green, J.A. and Prather, R.S.  
 TITLE Swine Genomics  
 JOURNAL Unpublished (2004)  
 COMMENT Contact: DNA Core Facility (Swine Project)  
 Animal Science - RS Prather  
 University of Missouri-Columbia  
 616 Medical Sciences Bldg., Columbia, MO 65212, USA  
 Tel: (573) 882-0428  
 Fax: (573) 884-5552  
 Email: porcine@net.missouri.edu  
 POLYA=yes.

FEATURES  
 source  
 1..235  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /dev\_stage="Blastocyst after cryopreservation"  
 /clone\_1lb="pcrtbcl1"  
 /note="Vector: GGTACACTA. Funding: A grant from the Monsanto Company to the University of Missouri. Genetic Source: Day 6 pig blastocyst stage embryos were flushed from a gilt's reproductive tract in TV Hapes and cultured in NCSU23 (1) for 3 hours prior to vitrification. For vitrification, embryos were cultured in 1M dimethylsulfoxide and 1.3M ethylene glycol in Tissue Culture Medium (TCM) and 20% newborn calf serum for 5 minutes at 39 degrees Celsius. Then embryos were transferred to 3.2M ethylene glycol and 0.6M sucrose in TCM and 20% newborn calf serum for 45 to 60 seconds at 39 degrees Celsius. Capillary action was used to load the embryos in open pulled straws and the straws were plunged into liquid nitrogen. At thawing embryos were placed in 0.2M sucrose in TCM and 20% newborn calf serum for 1 minute, then 0.1M sucrose in TCM and 20% newborn calf serum for 5 minutes. Embryos were then rinsed and cultured in NCSU23 for 12 or 24 hours. Library Construction (PCR Protocol): The amount of mRNA that was recovered from embryos was quite limiting and was not sufficient for library production with a standard protocol. Therefore, PCR-based protocol was utilized for producing libraries. The blastocysts were then rinsed with 50 ul polyvinylalcohol-PBS in dpcp treated medium three times and snap frozen with 1 ul super-RNasin in a 0.5ml RNase-free tube. Hybrid messenger affinity paper was used to isolated the mRNA (2). The mRNA was reverse transcribed with a NotI-tag-dT18 oligonucleotide and a SMART oligonucleotide (Clontech) modified to contain a SalI site to generate full-length cDNA with a sequence complementary to the SMART oligonucleotide. Sequences within the SMART and dT oligonucleotides were used as primers to amplify the cDNAs by PCR with pfu turbo polymerase (Stratagene). The resulting PCR products were purified, digested with NotI and SalI and size fractionated by using a Chroma Spin-400 followed by a Chroma Spin-1000 column (Clontech). Purified cDNA from each PCR reaction was ligated into the pCMV-SPORT6 vector. Preliminary library characterization: Randomly chosen clones from each library were analyzed by restriction digestion to determine average insert size (96 clones) and by sequencing (~3 96-well plates) to confirm library quality (e.g. the presence of short polyA tails,

## ORIGIN

Query Match 90.6%; Score 15.4; DB 7; Length 235;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCCGGTAGTACTT 17  
 |||||  
 Db 146 AGCCCGGTATGTACTT 162

genomic DNA contamination (must be <1%), ribosomal RNA clones (must be <1%), etc.) and to provide a sequence database representing the predominant clones in each library. The clones were sequenced at the University of Missouri-Columbia DNA Core Facility. Bioinformatics work was performed by GK Springer's bioinformatics group (MG Spollen, J Riese, A Gullin, AA Khambati, RV Patel, CM Topinka, SB Bhuiyan) in Computer Science and Health Management and Informatics Departments at the University of Missouri-Columbia. Clone Requests: Requests for clones should be made to the Director of the University of Missouri DNA Core Facility at: porcine@net.missouri.edu. Citations: 1. Im G-S, Lai L, Liu Z, Hao Y and Prather RS. In vitro development of preimplantation porcine nuclear transfer embryos cultured in different media and gas atmospheres. Theriogenology 61: 1125-1135, 2004. 2. Anderson JE, Matteri RL, Abeydeera LR, Day BN and Prather RS. Degradation of maternal Cdc25c during the maternal to zygotic transition is dependent upon embryonic transcription. Molecular Reproduction & Development 60: 181-188, 2001.  
 TAG TISSUE-Blastocyst after cryopreservation  
 TAG\_SEQ=GTGTAACCTA"

RESULT 3  
 BMS39228/c 411 bp mRNA linear EST 20-FEB-2002  
 LOCUS hb06f02.g1 Canis CDNA from testes cells Canis familiaris CDNA  
 DEFINITION clone hb06f02 5', mRNA sequence.  
 ACCESSION BMS39228  
 VERSION BMS39228.1 GI:18820982  
 KEYWORDS EST.  
 SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 1 (bases 1 to 411)  
 O'Shaughnessy, A.L., McCombie, W.R., Baker, J.P., Balija, V.,  
 Cummins, D., Dedhia, N.N., de la Bastide, M., Katzenberger, F.,  
 King, L., Kirchoff, K.A., Miller, B., Muller, S., Nascimento, L.U.,  
 Palmer, L., Santos, L., Shah, R.S., Spiegel, L.A., Zutavern, T.,  
 Preston, R. and Hannon, G.J.  
 REFERENCES  
 Expresseed sequence tags from Canis familiaris (dog) (2002)  
 Unpublished (2002)  
 CONTACT: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100 Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874  
 Email: mcombie@cshl.org  
 plate: hb06 row: f column: 02  
 Seq primer: -21M3univrev  
 High quality sequence stop: 411.

## FEATURES

source  
 1..411  
 /organism="Canis familiaris"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9615"  
 /clone="hb06f02"  
 /clone\_1lb="Canis CDNA from testes cells"  
 /note="Vector: lambda Zap II; The library was produced by

Greg Hannon and Raymond Preston (Cold Spring Harbor Laboratory). This library is oligo(dT) primed using stratagene zap cDNA synthesis kit. It was made from dog testes. Please contact Greg Hannon (hannon@cshl.org) with any library related inquiries."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 4; Length 411;  
Best Local Similarity 94.1%; Pred. No. 3.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17  
|||||  
22 AGCCTCGTAGTAACCT 6

Db

RESULT 4 CA217548/c 528 bp mRNA linear EST 25-SEP-2003  
LOCUS SCRUD1132D12.B ADI Saccharum officinarum cDNA clone SCRUD1132D12  
DEFINITION 3', mRNA sequence.  
ACCESSION CA217548  
VERSION CA217548.1 GI:35267956  
KEYWORDS EST.  
SOURCE Saccharum officinarum  
ORGANISM Saccharum officinarum

REFERENCE 1 (bases 1 to 528)  
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
The libraries that made SUCEST  
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
Contact: Arruda P  
Centro de Biologia Molecular e Engenharia Genetica  
Universidade Estadual de Campinas  
Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
Tel: 55 19 3788 1137  
Fax: 55 19 3788 1089  
Email: parruda@unicamp.br  
Clone distribution information can be found  
through the Brazilian Clone Collection Center (BCCC) at  
http://www.bcccenter.fcav.unesp.br  
Plate: 132 row: D column: 12  
Seg primer: SP6 Promoter primer.  
Location/Qualifiers

1. 528  
/organism="Saccharum officinarum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4547"  
/clone="SCRUD1132D12"  
/lab\_host="DH10B"  
/clone\_1lb="AD1"  
/note="Organ: seedlings inoculated with Gluconacetobacter diazotrophicans; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from seedlings inoculated with Gluconacetobacter diazotrophicans; cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://succest.lad.ic.unicamp.br/public"

## FEATURES

source

1. 528  
/organism="Saccharum officinarum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4547"  
/clone="SCRUD1132D12"  
/lab\_host="DH10B"  
/clone\_1lb="AD1"  
/note="Organ: seedlings inoculated with Gluconacetobacter diazotrophicans; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from seedlings inoculated with Gluconacetobacter diazotrophicans; cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://succest.lad.ic.unicamp.br/public"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 528;  
Best Local Similarity 94.1%; Pred. No. 3.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17  
|||||

Db 330 AGCCGCGTAGTAACCT 314

RESULT 5 CV224495/c 533 bp mRNA linear EST 21-SEP-2004  
LOCUS CS\_hyp\_19d10.M3Reverse Blue crab hypodermis, normalized  
DEFINITION Callinectes sapidus cDNA clone CS\_hyp\_19d10.5, similar to sp|P81385|CUB\_HOMAM Cuticle protein AMP1B (HA-AMP1B). Score = 43.9 bits (102), Expect = 0.002, mRNA sequence.  
CV224495  
CV224495.1 GI:52371724  
EST.  
VERSION CV224495.1 GI:52371724  
KEYWORDS EST.  
SOURCE Callinectes sapidus (blue crab)  
ORGANISM Callinectes sapidus (blue crab)

REFERENCE 1 (bases 1 to 533)  
Shafer,T.H., Coblentz,F.E. and Towle,D.W.  
Expressed sequence tags from normalized cDNA libraries prepared from gill and hypodermis tissues of the blue crab, Callinectes sapidus  
Unpublished (2004)  
Contact: Thomas H. Shafer  
Department of Biological Sciences  
University of North Carolina Wilmington  
601 S. College Rd, Wilmington, NC 28403, USA  
Tel: 910-962-7275  
Fax: 910-962-4066  
Email: shafer@uncw.edu  
Plate: 13 row: d column: 11  
Seg primer: M13 Reverse  
High quality sequence stop: 176.  
Location/Qualifiers

## JOURNAL

Unpublished (2004)  
Contact: Thomas H. Shafer  
Department of Biological Sciences  
University of North Carolina Wilmington  
601 S. College Rd, Wilmington, NC 28403, USA  
Tel: 910-962-7275  
Fax: 910-962-4066  
Email: shafer@uncw.edu  
Plate: 13 row: d column: 11  
Seg primer: M13 Reverse  
High quality sequence stop: 176.  
Location/Qualifiers

## FEATURES

source

1. 533  
/organism="Callinectes sapidus"  
/mol\_type="mRNA"  
/db\_xref="taxon:6763"  
/clone="CS\_hyp\_19d10"  
/issue\_type="Pooled hypodermal epithelium from the mid-dorsal region and antrochial membrane of premolt (stage D2) and 3-hour postmolt crabs"  
/dev\_stage="Adult"  
/clone\_1lb="Blue crab hypodermis, normalized"  
/note="Vector: pCMV Sport 6.1; Total RNA samples were prepared individually from each tissue, checked for quality, and then pooled for construction and normalization of a cDNA library by Invitrogen. Plasmids were isolated and inserts sequenced from their 5'-ends by the Blue Crab Molecular Genetics Laboratory at the University of North Carolina Wilmington. Traces were trimmed, compared (BLASTx) to NCBI non-redundant protein database as of 19 July 2004, and processed for submission to dbEST by trace2dbEST software (Parkinson, Anthony and Blaxter, unpublished software)."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 7; Length 533;  
Best Local Similarity 94.1%; Pred. No. 3.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17  
|||||  
469 AGCCACGAGTAGTAACCT 453

Db

RESULT 6 CV224494/c 667 bp mRNA linear EST 21-SEP-2004  
LOCUS CS\_hyp\_19d10.M3Reverse Blue crab hypodermis, normalized  
DEFINITION Callinectes sapidus cDNA clone CS\_hyp\_19d10.5, similar to sp|P81385|CUB\_HOMAM Cuticle protein AMP1B (HA-AMP1B). Score = 36.2

bites (82), Expect = 0.58, mRNA sequence.  
 CV224494  
 CV224494.1 GI:52371722  
 EST.  
 SOURCE Callinectes sapidus (blue crab)  
 ORGANISM Callinectes sapidus  
 Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 Eubrachyura; Portunoidae; Portunidae; Callinectes.  
 REFERENCE 1 (bases 1 to 667)  
 AUTHORS Shafer,T.H., Coblenz,F.B. and Towle,D.W.  
 TITLE Expressed sequence tags from normalized cDNA libraries prepared from gill and hypodermis tissues of the blue crab, Callinectes sapidus  
 JOURNAL Unpublished (2004)  
 COMMENT Contact: Thomas H. Shafer  
 Department of Biological Sciences  
 University of North Carolina Wilmington  
 601 S. College Rd. Wilmington, NC 28403, USA  
 Tel: 910-962-7275  
 Fax: 910-962-4066  
 Email: shafer@uncw.edu  
 Plate: 19 row: d column: 10  
 Seq primer: M13 Reverse  
 High quality sequence stop: 494.  
 FEATURES  
 source  
 1..667  
 Location/Qualifiers  
 /organism="Callinectes sapidus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:6763"  
 /clone="CS\_hyp\_19d10"  
 /tissue\_type="Pooled hypodermal epithelium from the mid-dorsal region and arthropodal membrane of premolt (stage D2) and 3-hour postmolt crabs"  
 /dev\_stage="Adult"  
 /clone\_1fb="Blue crab hypodermis, normalized"  
 /note="Vector: PCWY Sport 6.1; Total RNA samples were prepared individually, from each tissue, checked for quality, and then pooled for construction and normalization of a cDNA library by Invitrogen. Plasmids were isolated and inserts sequenced from their 5'-ends by the Blue Crab Molecular Genetics Laboratory at the University of North Carolina Wilmington. Traces were trimmed, compared (BLASTX) to NCBI non-redundant protein database as of 19 July 2004, and processed for submission to dbEST by trace2dbEST software (Parkinson, Anthony and Blaxter, unpublished software)."  
 ORIGIN  
 Query Match 90.6%; Score 15.4; DB 7; Length 667;  
 Best Local Similarity 94.1%; Pred. No. 3.9e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGCCGGGTAGTAACT 17  
 |||||  
 Db 468 AGCCAGTAACT 452  
 RESULT 7  
 LOCUS BQ294905/c 700 bp mRNA linear EST 16-MAY-2002  
 DEFINITION WHE2855.G07 M13S Wheat unexpressed root tip cDNA library Triticum aestivum cDNA clone WHE2855\_G07\_M13, mRNA sequence.  
 ACCESSION BQ294905  
 VERSION BQ294905.1 GI:20810125  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticaceae; Triticum.  
 REFERENCE 1 (bases 1 to 700)  
 AUTHORS Anderson,O.D., Chao,S., Chin,A., Close,T.J., Crossman,C.,

Gustafson,P., Lazo,G.R., Pham,J., Rauech,C.J., Ross,K., Wilson,C. and Woo,J.  
 TITLE The structure and function of the expressed portion of the wheat genomes - Unexpressed root tip cDNA library  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Olin Anderson  
 US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 510555773  
 Fax: 5105559518  
 Email: oanderson@wpr.usda.gov  
 Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
 Seq primer: SK primer  
 FEATURES  
 source  
 1..700  
 Location/Qualifiers  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"  
 /clone="WHE2855\_G07\_M13"  
 /tissue\_type="Root tip at 1.0 to 1.5 mm stage"  
 /dev\_stage="Four-day old seedling"  
 /lab\_host="E. coli SOLR"  
 /clone\_1fb="Wheat unexpressed root tip cDNA library"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
 Site\_1: EcoRI; Site\_2: XhoI; Plants were grown under hydroponic conditions for four days. Root tips were excised and snap frozen (Ross and Gustafson) and total RNA was prepared at University of Missouri, Columbia. Poly(A) RNA was purified, a cDNA library was made, and the cDNA clones were in vivo excised to give phluescript SK(-) phagemids in the T7 Close lab (Chin and Close) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."  
 ORIGIN  
 Query Match 90.6%; Score 15.4; DB 5; Length 700;  
 Best Local Similarity 94.1%; Pred. No. 3.9e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGCCGGGTAGTAACT 17  
 |||||  
 Db 24 AGCCGGGTAGTAACT 8  
 RESULT 8  
 LOCUS CL653433/c 768 bp DNA linear GSS 09-JUL-2004  
 DEFINITION PR10118c\_B08 - PR10118c.B21 (768) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.  
 ACCESSION CL653433  
 VERSION CL653433.1 GI:50132271  
 KEYWORDS GSS.  
 SOURCE Pristionchus pacificus  
 ORGANISM Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.  
 REFERENCE 1 (bases 1 to 768)  
 AUTHORS Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
 TITLE AppADB: An AcedB database for the nematode satellicite organism Pristionchus pacificus  
 JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)  
 COMMENT Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371  
 Fax: 00497071601498  
 Email: ralf.sommer@tuebingen.mpg.de



This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.  
Seq primer: T7  
Clase: fosmid end.

## FEATURES

source

Location/Qualifiers  
1..768  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Mixed stage fosmid library of P. pacificus  
var. California"  
/note="Vector: pBplfos-5 Fosmid vector"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 9; Length 768;  
Best Local Similarity 94.1%; Pred. No. 3.9e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACT 17  
|||||  
Db 542 AGCCGCGTAGTAACT 526

RESULT 9 509 bp mRNA linear EST 03-JUN-2003  
LOCUS CD445870  
DEFINITION EL01IT0204F06.b Endosperm\_4 Zea mays cDNA, mRNA sequence.  
ACCESSION CD445870  
VERSION CD445870.1 GI:31361513  
KEYWORDS EST.  
SOURCE Zea mays  
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 509)

Lai J., Dey, N., Kim, C.S., Bharti, A.K., Rudd, S., Mayer, K.F.,  
Larkin, B., Becraft, P. and Messing, J.

Characterization of the maize endosperm transcriptome and its  
comparison to the rice genome  
Genome Res. 14 (10), 1932-1937 (2004)

JOURNAL Contact: Lai, Jinheng  
COMMENT Dr. Joachim Messing's lab

Wakeman Institute, Rutgers University  
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
Tel: 732-445-3801  
Fax: 732-445-5735  
Email: jlai@wakeman.rutgers.edu

Seq primer: T3.  
FEATURES  
source

Location/Qualifiers  
1..509  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/rname\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm\_4"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; site\_2:  
XhoI"

## ORIGIN

Query Match 88.2%; Score 15; DB 6; Length 509;  
Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACT 15  
|||||  
Db 74 AGCCGCGTAGTAACT 88

RESULT 10  
CNS06U7P

LOCUS CNS06U7P 724 bp DNA linear GSS 06-JUN-2001  
DEFINITION T3 end of clone AX0AA005G05 of library AX0AA from strain CBS 7064  
of Pichia farinosa, genomic survey sequence.

ACCESSION AL415499  
VERSION AL415499.1 GI:12194769  
KEYWORDS GSS.  
SOURCE Pichia farinosa  
ORGANISM Pichia farinosa

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,  
1 (bases 1 to 724)  
Bolotin-Fukuhara, M., Bon, E., Brotier, P., Casaregola, S.,  
de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B.,  
Malpertuy, A., Nevegilise, C., Ozier-Kalogeropoulos, O., Potier, S.,  
Saurin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M.,  
Wincker, P. and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
FEMS Lett. 487 (1), 3-12 (2000)  
JOURNAL  
MEDLINE 20584711  
PUBMED 11152876

REFERENCE 2 (bases 1 to 724)  
de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B.,  
Wincker, P., Artiguenave, F. and Potier, S.

Genomic exploration of the hemiascomycetous yeasts: 15. Pichia  
sorbitophila  
FEMS Lett. 487 (1), 87-90 (2000)

JOURNAL  
MEDLINE 20584725  
PUBMED 11152890

REFERENCE 3 (bases 1 to 724)  
Genoscope.

Direct Submission  
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail :  
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

## COMMENT

## FEATURES

source

Location/Qualifiers  
1..724  
/organism="Pichia farinosa"  
/mol\_type="genomic DNA"  
/strain="CBS 7064"  
/db\_xref="taxon:4920"  
/clone="AX0AA005G05"  
/clone\_lib="AX0AA"  
/note="end : T3"

## ORIGIN

Query Match 88.2%; Score 15; DB 9; Length 724;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCGCGTAGTAACT 17  
|||||  
Db 207 CCGCGTAGTAACT 221

RESULT 11 821 bp DNA linear GSS 17-MAR-2003  
BZ797610  
LOCUS BZ797610/c  
DEFINITION PUFCT697D\_ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMBFaz91K17,  
genomic survey sequence.  
ACCESSION BZ797610  
VERSION BZ797610.1 GI:29000148

```

KEYWORDS      GSS.
SOURCE         Zea mays
ORGANISM       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
                clade; Panicoidae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 821)
                Whiteaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
                Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
                Bennetzen,J.
TITLE          Maize Genomics Consortium
JOURNAL        Unpublished (2003)
COMMENT        Other_GSSes: PUCY69TB
                Contact: Cathy Whiteaw
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA
                Tel: 301-838-5843
                Fax: 301-838-0208
                Email: whiteaw@tigr.org
                Seq primer: TF
                Class: sheared ends.
FEATURES       Location/Qualifiers
                1..821
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /strain="B73"
                /db_xref="taxon:4577"
                /clone="ZMBR077612"
                /clone_1b="ZM 0.6_1.0 KB"
                /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
                Cor selected genomic DNA library"

ORIGIN
Query Match      88.2%; Score 15; DB 8; Length 821;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGCGTAGGTAC 15
        |||||||
        323 AGCCGCGTAGGTAC 309

RESULT 12      927 bp      DNA      linear      GSS 19-AUG-2003
CG063514      PUKAR59TD_ZM_0.6_1.0_KB Zea mays genomic clone ZMBR0776122,
LOCUS          genomic survey sequence.
DEFINITION     CG063514
ACCESSION      CG063514
VERSION        CG063514.1 GI:33935694
KEYWORDS       GSS.
SOURCE         Zea mays
ORGANISM       Zea mays
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
                clade; Panicoidae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 927)
                Whiteaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
                Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
                Bennetzen,J.
TITLE          Maize Genomics Consortium
JOURNAL        Unpublished (2003)
COMMENT        Other_GSSes: PUKAR59TB
                Contact: Cathy Whiteaw
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA
                Tel: 301-838-5843
                Fax: 301-838-0208
                Email: whiteaw@tigr.org
                Seq primer: TF
                Class: sheared ends.
FEATURES       Location/Qualifiers
                1..927
                /organism="Zea mays"

```

```

/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMBR0776122"
/clone_1b="ZM 0.6_1.0 KB"
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
Cor selected genomic DNA library"

ORIGIN
Query Match      88.2%; Score 15; DB 9; Length 927;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGCGTAGGTAC 15
        |||||||
        737 AGCCGCGTAGGTAC 723

RESULT 13      1107 bp      mRNA      linear      EST 21-DEC-2000
BF672870      602152718r1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293820 5',
LOCUS          mRNA sequence.
DEFINITION     BF672870.1 GI:11946765
ACCESSION      BF672870
VERSION        BF672870.1
KEYWORDS       Homo sapiens (human)
SOURCE         Homo sapiens
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 1107)
                NIH-MGC http://mgs.nci.nih.gov/.
                National Institutes of Health, Mammalian Gene Collection (MGC)
                Unpublished (1999)
                Contact: Robert Strausberg, Ph.D.
                Email: cga@bbs-rtm1.nhl.nih.gov
                Tissue Procurement: CLONETECH Laboratories, Inc.
                cDNA Library Preparation: CLONETECH Laboratories, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                plate: L1CML142 row: C column: 05
                High quality sequence start: 33
                High quality sequence stop: 49.
FEATURES       Location/Qualifiers
                1..1107
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4293820"
                /lab_host="DH10B (T1 phage-resistant)"
                /clone_1b="NIH_MGC_81"
                /note="Organ: muscle (skeletal); Vector: pDNR-LIB
                (Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI
                (ggcgctatggcc); 5' and 3' adaptor were used in cloning
                as follows: 5'-CACGCGCATATGCGC-3'
                and 3' adaptor sequence:
                5'-ATTCTAGAGCGGAGCGGCGGACATG-dt(30)BN-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size
                1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
                inserts by PCR. This library was enriched for full-length
                clones and was constructed by Clontech Laboratories (Palo
                Alto, CA)."

```

```
COMMENT      Contact: Ben C  
             Laboratoire BAP  
             Inpt - Ensac, Pole de Biotechnologie Vegetale  
             18 Chemin de Bode Rouge, Anzeville, Castanet Tolosan 31326,  
             FRANCE.  
  
FEATURES  
source      Location/Qualifiers  
            1..419  
              /organism="Helianthus annuus"  
              /mol_type="mRNA"  
              /db_xref="taxon:4232"  
              /clone="HacoCotR005B01"  
              /dev_stage="cotyledonary embryo"  
              /clone_1lb="HacoCR"  
              /note="Helianthus annuus cotyledonary embryo; inbred line:  
                Ha300b"
```

2.3 Egg Bank (uspiog)